VOLUME 23, 1961/62, No. 5 pp. 133-160 Published 7th February 1962

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

AN EXPERIMENTAL FLUORESCENT SCREEN IN DIRECT-VIEWING TUBES FOR COLOUR TELEVISION

by R. R. BATHELT *) and G. A. W. VERMEULEN *).

535.371.07:621.397.62:621.397.132

The screen of direct-viewing tubes for colour television consists of three phosphors, giving red, green and blue fluorescence. The phosphors hitherto used have been, respectively, a phosphate, a silicate and a sulphide. A serious drawback of the phosphate and silicate is their relatively long afterglow. Red and green fluorescent phosphors are known that have a much shorter afterglow, but their use has been barred by the difficulty of applying them to the face plate. Methods of overcoming this difficulty have been studied at Philips since 1957. The technique described in this article, although still in the experimental stage, offers good prospects of direct-viewing screens composed of three sulphide phosphors.

The principles of colour television have already been dealt with in this journal [1]). Briefly, the camera produces three images of the scene in the primary colours red, green and blue, and the three video signals thus obtained (R, G and B respectively), after being coded, are transmitted modulated on a carrier; after decoding in the receiver, three corresponding video signals R, G and B are available, and it is the task of the display device to modulate a red, a green and a blue light source in accordance with these signals.

A system developed for this purpose is the *colour-television projector* earlier described [2]). This uses three projection tubes having phosphors which fluoresce red, green and blue respectively. The tubes are separately driven by the signals R, G and B, producing a red, a green and a blue image. The three images are magnified by an optical system and projected on to a screen in such a way that they accurately coincide.

Although this system can also be adopted for smaller pictures, *direct-viewing colour picture tubes* are preferable for use in the home.

Direct-viewing colour picture tubes

Several types of these tubes exist. They are all designed with the common object of producing a complete colour picture on the screen by means of the signals R, G and B. They might be said to be a combination of three ordinary picture tubes — but each having a screen which fluoresces in a different primary colour — in a single envelope.

As in all cathode-ray tubes, the fluorescent screen is a coating on the inside surface of the tube face. It consists of a regular array of large numbers of colour cells (several hundred thousand in the shadow-mask tube to be mentioned later). Each colour cell contains a specific geometrical arrangement of three primary phosphors, which fluoresce red, green and blue, respectively, upon bombardment by electrons. The arrangement of these phosphors may differ widely: they may be applied as closely packed parallel lines in the sequence red-green-blue-red-green-blue-etc., but usually they are dots with the red, green and blue of each triad contiguous to one another. The shortest distance between two dots of the same colour must be small enough for the structure to be indistinguishable by the eye at the normal viewing distance, where it is thus seen only as a mixture of the light emitted by the individual phosphors (cf. colour printing).

---

*) Electron Tubes Division, Eindhoven.

[1]) F. W. de Vrijer, Fundamentals of colour television, Philips tech. Rev. **19**, 86-97, 1957/58.

[2]) T. Poorter and F. W. de Vrijer, The projection of colour-television pictures, Philips tech. Rev. **19**, 338-355, 1957/58.

The phosphors in each cell must be bombarded by electrons in accordance with the instantaneous values of the three primary signals. This can be done in two ways:

1) *simultaneously*, by means of three electron guns (one for each primary colour) which are mounted side by side in the neck of the tube, or
2) *sequentially*, by means of one electron gun, the beam striking in turn a red, a green and a blue phosphor. In this method electronic means are needed to ensure that the gun is driven at every instant by the appropriate signal.

Further, the tube must also contain a colour-selecting device, which directs the electrons on to the appropriate primary phosphors. It is particularly in this respect that the various types of colour tubes differ one from the other. We shall not go into these differences here; later on in this article, one particular type — the shadow-mask tube — will be dealt with at greater length.

Conventional method of applying the phosphors

For applying the three phosphors in a specific pattern to the inside face of the picture tube, a kind of photographic process is nowadays employed. Use is made of a lacquer which is polymerized and hardened under ultraviolet irradiation. The lacquer is usually a solution of polyvinyl alcohol in water, sensitized with a dichromate. The maximum sensitivity lies at a wavelength of 365 nm, i.e. in the long-wave ultraviolet; the boundary wavelength is in the blue, in the region of 470 nm (1 nm = 10^{-9} m) (*fig. 1*).

The first phosphor to be applied to the face plate is suspended in the lacquer and the resultant slurry is coated uniformly over the tube face. After drying, the coating is irradiated only at those places where the phosphor is required. Depending on the type of tube, this is done by producing an optical image of a particular negative on the face plate by means of ultraviolet radiation, or by the shadow effect obtained by irradiating a negative with a point source of ultraviolet radiation; these negatives contain the appropriate pattern of lines or dots. The lacquer is hardened only at the places irradiated, so that a latent image is produced which is developed in the next operation. This operation is a water treatment, the unwanted phosphor being removed by causing the lacquer at the non-irradiated areas to swell up and dissolve. At the irradiated places the lacquer is insoluble and remains adhering to the glass together with the phosphor. The result, after drying, is the first phosphor pattern, still mixed with the polyvinyl alcohol.

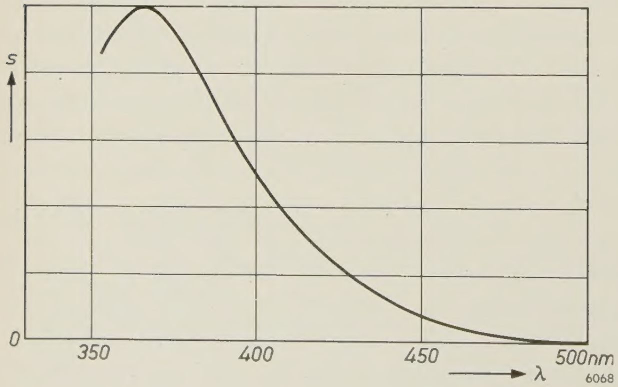


Fig. 1. Spectral sensitivity characteristic of the hardening of a polyvinyl-alcohol solution sensitized with dichromate. The maximum lies in the long-wave ultraviolet at 365 nm. (1 nm = 10^{-9} m.)

The process is now repeated with the second phosphor, and then again with the third phosphor. The irradiation must be done with other negatives or from other points, since of course the three patterns must be displaced with respect to one another.

Once the fluorescent screen is completed, it is aluminized in the same way as in black-and-white tubes, being coated with some organic material (generally a methacrylate) on which a film of aluminium is then vacuum-deposited. The aluminium backing serves, among other things, as a reflector. Finally the entire screen is baked out to remove the polyvinyl alcohol and the methacrylate.

Choice of phosphors

The choice of the primary colours was dealt with at some length in the article quoted [3]). Apart from the colour of the emitted light and the luminous efficiency, important considerations in the choice of phosphors are the afterglow (persistence) and the technological properties of the phosphors in regard to their application in colour tubes. Afterglow in moving television pictures must not be too long, for if the luminance of one field (i.e. of one of the two equal parts into which the picture is divided in the interlaced scanning method normally used in television) is too high after 20 milliseconds, when the next field is being traced, the definition will be reduced, and moreover "trailing" will be noticeable at marked transitions of brightness.

The following are the main phosphors hitherto used in direct-viewing colour picture tubes:

for red, a zinc phosphate activated with manganese;
for green, willemite (a zinc silicate) activated with manganese;
for blue, a zinc sulphide activated with silver.

The colours in which these phosphors fluoresce are the primary colours of the colour television system adopted in the United States — the N.T.S.C. system

---

[3]) See ref. [1]), p. 90, fig. 8.

(named after the National Television System Committee)[4].

Table I gives the major properties of the above-mentioned phosphors, and for comparison the data of the normal phosphors used in black-and-white tubes (a mixture of yellow and blue). As appears from the table, the red and green phosphors have a much higher persistence than the blue phosphor, too high in fact. This causes the above-mentioned loss of definition in moving images and also, depending on the colours in the scene, gives rise to "trailing" — streaks of red, orange or green hue behind sharp transitions.

Red and green fluorescent phosphors are known, however, that have a lower persistence. As a green phosphor, for example, willemite can be used with a greater activator content. The persistence is then much lower, although still not as low as might be wished; moreover the luminous efficiency is somewhat lower.

A better solution is offered by the zinc-cadmium sulphides having a suitably chosen cadmium-sulphide content. With increasing cadmium-sulphide content the colour of the light emitted by this phosphor changes from blue through green, yellow and orange to red. The dot-dash line in the chromaticity diagram shown in *fig. 2* gives the colour points for the fluorescence of zinc-cadmium sulphides, the cadmium-sulphide content being indicated at some points along the curve. A yellow-fluorescent (ZnCd)S, as mentioned in table I, is used as the yellow component in black-and-white tubes. All (ZnCd)S phosphors have a low persistence.

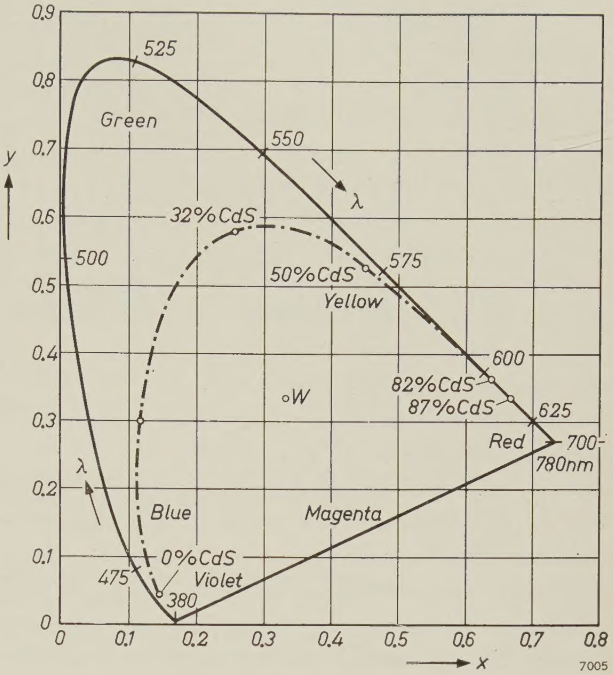


Fig. 2. Chromaticity diagram. The dot-dash line gives the colour points of the fluorescent light from (ZnCd)S phosphors with varying cadmium-sulphide content. *W* is the colour point of white.

In addition to a low persistence the zinc-cadmium sulphides have the advantage of a high efficiency. *Table II* gives the data for the strongly activated willemite just mentioned and for the zinc-cadmium sulphides suitable for green and red. With an appropriately chosen cadmium-sulphide content it is possible to give red-fluorescent (ZnCd)S the same colour point as the $Zn_3(PO_4)_2$-Mn used hitherto. The green of (ZnCd)S is less saturated than that of willemite; we shall return to this point at the end of the article.

In view of the high efficiency and low persistence of the green-fluorescent and red-fluorescent sulphide, it might be asked why these long familiar phosphors

Table I. Data of conventional phosphors in colour picture tubes and black-and-white tubes.

Type of tube	Colour	Phosphor	Colour co-ordinates		Efficiency *) cd/W	Persistence **) %
			x	y		
colour tube	red green blue	$Zn_3(PO_4)_2$-Mn Zn_2SiO_4-Mn ZnS-Ag	0.670 0.210 0.140	0.330 0.710 0.080	1.5 5.5 1.4-2.3	29 18 <1
black- and- white tube	yellow blue	(ZnCd)S-Ag ZnS-Ag	0.450 0.145	0.530 0.100	6.5	<1 <1

*) Luminous intensity in a direction perpendicular to an aluminized fluorescent screen radiating in accordance with Lambert's law, per watt of electric power supplied. The actual efficiencies of the phosphors are in fact higher: the figures given here allow for the 72% transmission of the tube face.

**) Luminance 20 milliseconds after electron-beam cut-off, as a percentage of the initial value.

---

[4] On direct-viewing tubes the resultant colour of the three primary rasters will be somewhat less saturated than the colours of the individual phosphors, owing to the fact that some of the fast secondary electrons liberated upon the bombardment of one of the three phosphors strike the two other phosphors.

were not used right from the beginning in colour picture tubes. The reason is that the red-fluorescent (ZnCd)S cannot be applied to the face plate by the method described above, whereby the lacquer containing the phosphor is irradiated with ultraviolet from the side where the electron gun will later be fitted. The zinc-cadmium sulphides get a deeper yellow colour as the cadmium-sulphide content increases, and strongly absorb ultraviolet, violet and blue. Upon irradiation a layer of the necessary thickness would therefore not be thoroughly hardened right up to the glass, and would flake off to a greater or lesser extent during the development process. The green-fluorescent sulphide is so weakly yellow that it can still just be applied in the normal manner, but with increasing cadmium-sulphide content the hardening radiation is so strongly absorbed that the method can no longer be used for the red-fluorescent sulphide.

Table II. Data of strongly activated willemite and of green- and red-fluorescent zinc-cadmium sulphide.

Colour	Phosphor	Colour co-ordinates		Efficiency *) cd/W	Persistence **) %
		x	y		
green	Zn_2SiO_4-Mn, strongly activated	0.250	0.700	4.5	4.5
green	(ZnCd)S-Ag 32% CdS	0.255	0.575	8.4	1-1.5
red	(ZnCd)S-Ag 87% CdS	0.670	0.330	1.8	<1
	(ZnCd)S-Ag 82% CdS	0.640	0.360	2.3	<1

*) and **): see Table I.

Efforts have been made to find other methods of applying the sulphide phosphors. In principle, ultraviolet-absorbent phosphors can be applied photographically by the "sticky-dots" method [5], in which the phosphor is made to stick to a previously applied lacquer pattern. The necessary stickiness is obtained by irradiating the lacquer for so short a time that it does not harden completely.

In this way, however, only an extremely thin layer of phosphor can be obtained. It appears to be difficult to prevent holes forming in the layer, as a result of which the luminous intensity is both inadequate and non-uniform over the screen. The whole procedure, starting from the application of the lacquer coating and ending with the development and drying of the phosphor layer, would therefore have to be repeated. In our experience, however, the result is even then unsatisfactory, and moreover it is extremely difficult to meet the requirements in regard to the average size of the grains and the grain-size distribution of the phosphors.

A new method of applying sulphide phosphors

In a new method developed here, use is made of the circumstance that it is not necessary to fix all three phosphors with the aid of a negative, but only two of them; the spaces left open by these two can be filled with the third phosphor, which is fixed by irradiation through the glass (without a negative), i.e. from outside the tube. Since the ultraviolet now has its maximum intensity at the interface between glass and lacquer, powerful adhesion to the glass is obtained.

The last phosphor to be applied (the red-fluorescent one) will now however be not only between but also (seen from outside) behind the green and the blue phosphors, and as a result the green and in particular the blue fluorescent light will be contaminated by red. Steps are therefore needed to prevent the red-fluorescent phosphor from being deposited behind the other phosphors. For this purpose, before the lacquer suspension containing red-fluorescent phosphor is applied, the green-fluorescent and blue-fluorescent phosphors are temporarily coated with an ultraviolet-absorbent dye on the gun side. At these positions, then, the lacquer with the red-fluorescent phosphor does not harden upon irradiation through the glass and is therefore removed in the development process.

Fig. 3 illustrates schematically in column *a* the method of fixing a red-fluorescent phosphate, which absorbs virtually no ultraviolet, and in columns *b* and *c* the fixing of a red-fluorescent sulphide that does absorb ultraviolet, respectively without and with a dye on the other phosphors. In case *b* the red-fluorescent phosphor is also deposited behind the other phosphors; in case *c* this is prevented by the protective dye.

After the last phosphor has been applied, the dye coating is dissolved.

Application of the dye

The dye can be applied in very much the same way as the phosphor. The ultraviolet-absorbent dye is

---

[5] M. Sadowski and P. D. Payne Jr., Photodeposition of luminescent screens, J. Electrochem. Soc. **105**, 105-107, 1958.

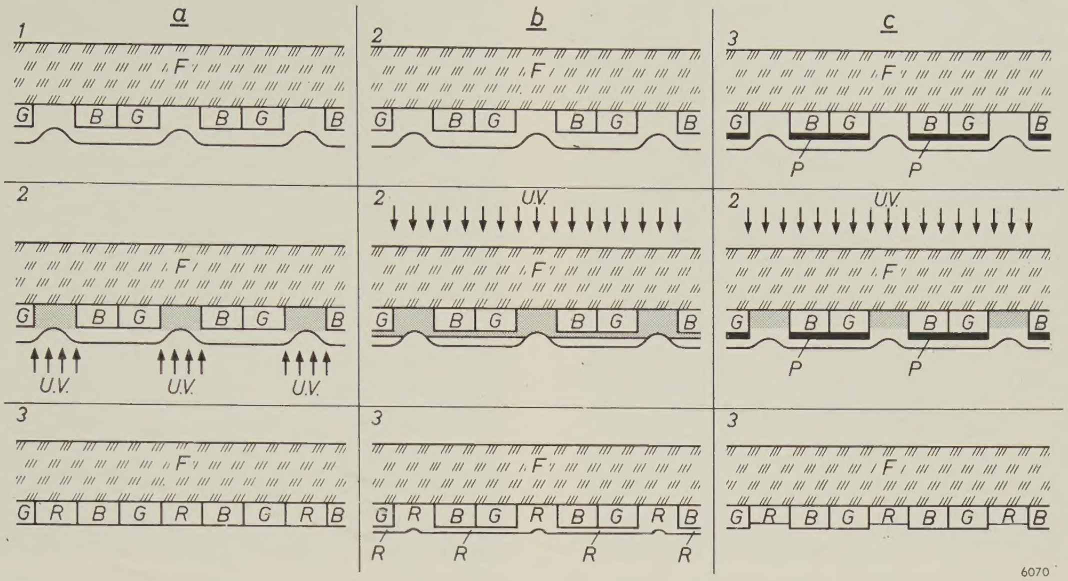


Fig. 3. Sketch representing schematically the application of the red-fluorescent phosphor (R) between the green-fluorescent and blue-fluorescent phosphor patterns (G and B).

Column a: The conventional method (red-fluorescent phosphor absorbing little ultraviolet).
1) Tube face F with G and B phosphor patterns, the latter being covered with a lacquer containing the R phosphor in suspension.
2) The places where the R phosphor dots are to come are irradiated from the gun side with ultraviolet. The shading indicates where the lacquer hardens.
3) Upon development the unhardened lacquer is removed, leaving behind the R phosphor dots.

Column b: The R phosphor is now a sulphide and absorbs ultraviolet. No dye is yet applied to the other phosphors.
1) As a, 1) above.
2) Irradiation with ultraviolet through the tube face. The shading indicates where the lacquer hardens.
3) After development the R phosphor remains, not only between but also (seen from outside) *behind* the G and B phosphors. As a result, the green and blue emissions are contaminated with red.

Column c: The new method.
1) The G and B phosphors are coated with a dye P which absorbs ultraviolet.
2) Hardening by irradiation through the glass.
3) After development the R phosphor remains only *between* the G and B phosphors.

suspended in a solution of polyvinyl alcohol sensitized with dichromate, and a coating of this suspension is applied to the screen, to which the green-fluorescent and blue-fluorescent patterns have already been fixed; this coating is then irradiated in such a way that both patterns are entirely covered with the dye after development.

For the purpose of analysing the principle of this method we shall assume first of all that the dye and afterwards the red-fluorescent phosphor are fixed with radiation of the same spectral composition (effective range from about 450 to 350 nm) and that the dye does *not* absorb selectively in this range.

The curve in *fig. 4* illustrates qualitatively the decrease in the intensity I of the radiation within the layer to be hardened, as a function of the depth of penetration d. At $d = d_p$ the intensity I has dropped to the minimum value I_{min} which, at a given irradiation time T_p, is needed to produce hardening. What remains after development is thus a layer of dye of thickness d_p. The transmittance of this layer

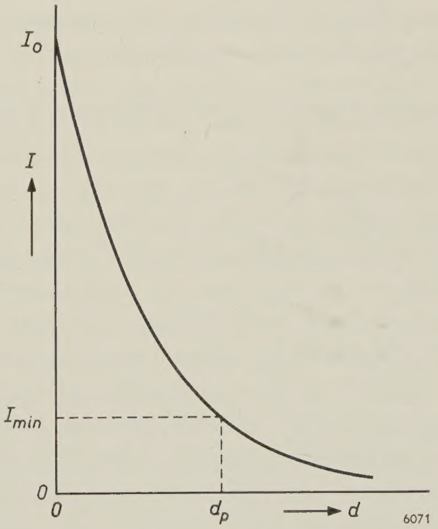


Fig. 4. Approximate variation of the intensity I of ultraviolet as a function of the depth of penetration d in a lacquer coating containing a dye in suspension. At $d = d_p$ the intensity has dropped to the minimum value I_{min} which, at a given exposure time T_p, is necessary to harden the lacquer. The result after development is a layer of dye of thickness d_p and transmittance I_{min}/I_0.

is not zero but equal to $I_{\min}/I_0$, where I_0 is the intensity at $d = 0$. When the suspension with the red-fluorescent phosphor has been applied and is irradiated through the glass, this radiation reaches the layer where it joins the glass with the intensity $a_g I_0$, and where the layer is separated from the glass by the green-fluorescent or blue-fluorescent phosphor, plus the dye, with the lower intensity $a_g a_f a_p I_0$. (The coefficients a allow for the loss by reflexion at an interface and by absorption in a medium; the suffixes g, f and p relate to glass, phosphor and dye respectively.) In order to fix the red-fluorescent phosphor solely at the places first mentioned, the irradiation time T_r must be so chosen that the exposure dose $a_g I_0 T_r$ causes hardening, but the smaller dose $a_g a_f a_p I_0 T_r$ does not. The result is therefore markedly dependent on T_r and on the factor a_p. Now a_p depends on the dye thickness d_p, and hence on the irradiation time T_p needed to fix the dye. It therefore depends critically on the ratio between the irradiation times T_p and T_r.

The value of this ratio becomes less critical if we choose a dye that absorbs selectively, such that a_p is small at the wavelength 365 nm to which the lacquer is most sensitive (see fig. 1), and much larger at longer wavelengths. For fixing the dye this makes it possible to use radiation mainly consisting of longer waves (from a different source) than the radiation used for fixing the red-fluorescent phosphor.

It should be noted that the dye is indispensable on the blue-fluorescent phosphor dots, but if necessary may be dispensed with on the green-fluorescent dots, provided the latter are composed of a somewhat yellowish (ZnCd)S; this phosphor itself absorbs the ultraviolet to a sufficient extent.

The new method applied to the shadow-mask tube

The new method is suitable for any type of direct-viewing colour picture tube. We shall presently discuss its application in the shadow-mask tube [6], but first it will be useful to give a brief description of this tube.

The shadow-mask tube contains three electron guns and a fluorescent screen with the primary phosphor applied as dots arranged in a hexagonal pattern (*fig. 5*). Each colour cell consists of one red, one green and one blue phosphor dot. The centres of these dots lie on the corners of equilateral triangles. In the optimum arrangement the dots touch one another without overlapping. The open spaces between the dots are later covered by the aluminium backing.

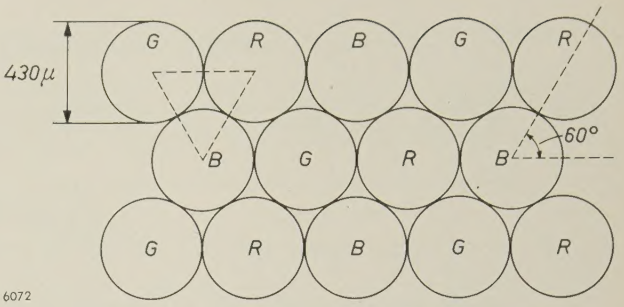


Fig. 5. Hexagonal configuration of the phosphor dots in the fluorescent screen of a shadow-mask tube. Each triad of phosphor dots fluorescing red, green and blue (*R*, *G*, *B*) forms a colour cell.

The three guns are mounted side by side in the neck of the tube. Colour selection is effected with the *shadow mask*, which is a perforated metal plate fitted in the tube about 13 mm behind the screen (*fig. 6*). Each colour cell has a corresponding hole in the shadow mask, which means that there are three times as many phosphor dots on the screen as there are holes in the mask. The guns are tilted slightly, so that their axes intersect on the shadow mask. The phosphor dots are so disposed on the tube face that the beam from each gun, after being deflected in the deflection coils, can pass through the hole in the mask only on to the correct phosphor. Electrons from the "red" gun, for example, strike only red-fluorescent dots, electrons from the "blue" gun strike only blue-fluorescent dots, and so on. Colour selection is therefore effected solely by masking.

For applying the phosphor dots by the photographic method no negative is needed during exposure in this case, since the shadow mask itself can serve as a "jig". Beyond the bend which they undergo in the deflection coils the electron trajectories from one

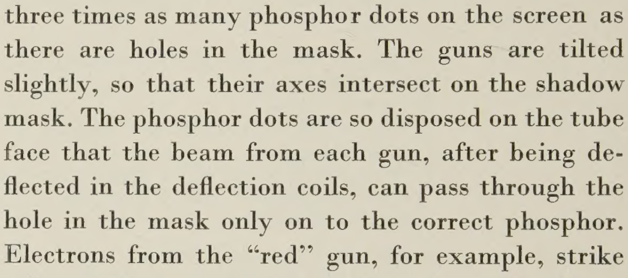
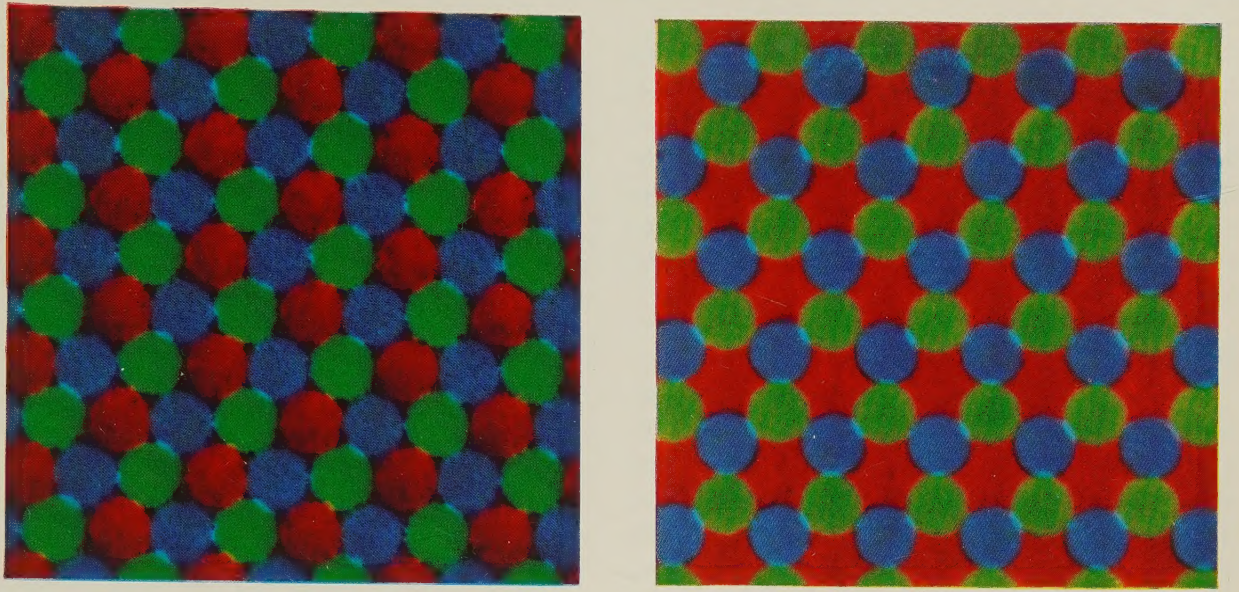


Fig. 6. Illustrating the principle of the shadow-mask tube. *F* tube face with hexagonal phosphor-dot pattern *R-G-B* as in fig. 5. *M* shadow mask. *r, g, b* electron beams from the "red", "green" and "blue" guns respectively. The configuration is such that each of the three beams strikes only phosphor dots of the corresponding colour.

---

[6] H. B. Law, A three-gun shadow-mask color kinescope, Proc. Inst. Radio Engrs. **39**, 1186-1194, 1951.
N. F. Fyler, W. E. Rowe and C. W. Cain, The CBS-Colortron: a color-picture tube of advanced design, Proc. Inst. Radio Engrs. **42**, 326-334, 1954.



a *b*

Fig. 7. Colour photographs of the screen of shadow-mask tubes. *a*) Screen with normal phosphors. *b*) Screen with three sulphide phosphors applied by the new method.
To make the structure of the screen visible, the guns were strongly defocused. In correct focus the electrons strike only parts of the phosphor dots that have a diameter of about 0.8 times the dot diameter.

gun in the tube are straight lines which, continued backwards, intersect at one point, the deflection point. The rays from an ultraviolet source situated at the deflection point thus pass through the shadow mask and strike those places on the screen where the phosphor dots of the relevant colour have to come, and only there do they harden the lacquer. In the successive irradiations the source must of course be situated at the deflection point for the primary colour in question.

In its present version 7) the shadow-mask tube is a round glass tube having a deflection angle of 70° and a convex screen of 53 cm diameter. The screen has roughly 350 000 colour cells. The holes in the shadow mask are made as large as is consistent with the requirement that, under normal conditions, the electrons are only just prevented from striking phosphor dots of the wrong colour. Nevertheless, a large proportion of the electron current from each gun is lost in the mask itself, where it generates heat; the average transmittance of the mask is only 15%. *Fig. 7a* reproduces a 20-times enlarged colour photograph of the colour screen. To make the entire screen structure visible, the guns were strongly defocused (normally the electrons strike only parts of the phosphor dots that have a diameter of about 0.8 times the total diameter of the dot).

In order to provide such a tube with three sulphide phosphors by the new method, the first procedure is to apply the green-fluorescent and blue-fluorescent phosphor dots in the normal way. Next, the dye suspension is similarly applied, after which it is dried and irradiated through the shadow mask in the same manner as for the blue-fluorescent phosphor. A yellow layer of dye forms behind each blue-fluorescent dot upon development. The screen is now coated with lacquer containing the red-fluorescent phosphor in suspension, and the lacquer is irradiated through the tube face. During this exposure, in which the shadow mask is not used, the lacquer hardens everywhere except behind the blue-fluorescent dots (owing to absorption in the dye) and behind the green-fluorescent dots (which are sufficiently absorbent themselves). The result after development is that all open spaces between the green-fluorescent and the blue-fluorescent pattern are covered with red-fluorescent phosphor. This can be seen from the photograph in fig. 7*b*, taken under similar conditions as for fig. 7*a* 8).

An important advantage of the new method in tubes with phosphor dots (as opposed to those with phosphor lines) is the following. The phosphor-dot

---

7) C. P. Smith, A. M. Morrell and R. C. Demmy, Design and development of the 21CYP22 21-inch glass color picture tube, RCA Rev. **19**, 334-348, 1958.

8) During the preparation of this article it came to the authors' attention that the Radio Corporation of America had brought out a shadow-mask tube using three sulphide phosphors. The fluorescent pattern is a three-dot configuration, which means that the red dots are not applied in the manner described here.

pattern of a normal colour picture tube (fig. 5) shows open spaces between the dots. Through these openings light from lamps in the room may be reflected by the aluminium backing. In the new method the openings between the green and blue phosphors are completely filled by the red phosphor, and the aluminium backing is thus unable to cause any troublesome reflections.

Tubes with two and three sulphides

Tubes using a zinc-cadmium sulphide for red, zinc sulphide for blue, and strongly activated willemite — with its relatively low persistence — for green, have not proved to be entirely satisfactory. Although the persistence is appreciably lower than in a tube using normal phosphors, so that moving images are sharper, the dark green fringes that appear during movement (owing to the fact that willemite still has a longer persistence than the sulphides, see Table II) have proved to be disadvantageous. Another drawback is that willemite has a lower efficiency than phosphors of the sulphide type.

Using three sulphides, i.e. a sulphide also for the green, gives the advantages of excellent definition for moving images and a higher luminance. Compared with the normal phosphors the gain in luminance is 20 to about 53% in the red (depending on the colour co-ordinates), 49% in the green and about 43% in the white ($x = 0.287, y = 0.316$).

Set against these substantial advantages is the drawback that green-fluorescent (ZnCd)S gives a less saturated colour than willemite. Generally speaking, where one of the primary colours (in this case green) has a lower saturation the result is a less faithful reproduction in two respects:

a) It narrows the range of colours reproducible in the picture.

b) It gives rise to colour errors, owing to the fact that the N.T.S.C. system hitherto used for colour television takes the fluorescence of willemite as the basis for green.

To conclude, we shall examine these points in more detail.

The dependence of the reproducible range of colours on the colour point of the green

The chromaticity diagram shown in *fig. 8* indicates the colour points R, G and B of the three normal phosphors. As will be known, all colours that can be produced by additive mixing of these three primaries lie within and on the periphery *1* of the triangle RGB.

A tube using sulphide phosphors has the same primary blue. Both for red and green there is a choice from a series of phosphors of widely differing colours. In fig. 8 a red-fluorescent sulphide phosphor has been chosen (with 87% CdS) whose colour point coincides with R. For green we have taken a (ZnCd)S with 32% CdS, which has roughly the same dominant wavelength as willemite. The colour point of this sulphide phosphor is denoted by G_s. Contour *2* thus bounds all colours that can be reproduced with these three sulphide phosphors. The contour *3* bounds all reflection colours of natural and artificial dyes and printing inks [9]).

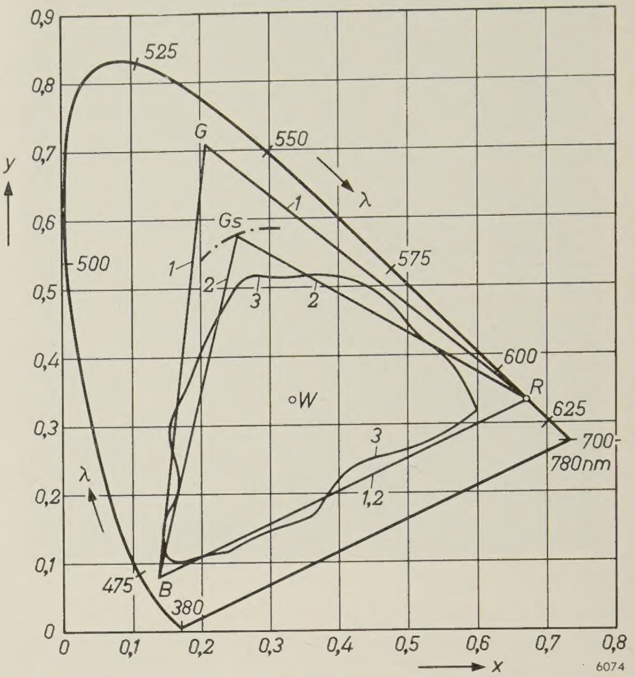


Fig. 8. Chromaticity diagram with the colour points R, G and B of the normal phosphors of direct-viewing colour picture tubes. All colours reproducible with these phosphors lie within or on the periphery *1* of triangle RGB. A sulphide phosphor with 87% CdS has its colour point also at R, a sulphide phosphor with 32% CdS at G_s. A screen containing both these sulphides and having the normal sulphide for blue can produce all colours lying within or on the periphery *2* of triangle RG_sB. Contour *3* marks the boundary of all reflection colours of the natural and artificial dyes. W is the white point.

With increasing cadmium-sulphide content in the green-fluorescent sulphide, G_s moves to the right along the dot-dash line.

It can be seen from fig. 8 that nearly as many colours can be reproduced with the three sulphide phosphors as with the conventional phosphors. The difference, which is due to the lower saturation of the green, is in reality even smaller than the figure might lead one to suppose, for as far as the green hues are concerned the change-over from G to G_s implies no loss in colour reproduction, since G_s is still outside the contour *3* of the colours which actually occur. (Even in the greenest parts of the picture, therefore, purely green fluorescence will never be present, but these parts will also contain red and blue.) There is, however, a slight loss in the reproduction of orange, yellow and greenish blue. The loss in yellow is smaller than might be inferred from fig. 8, because of the fact that the difference between colours whose chromaticity points are the same distances apart are much less noticeable in yellow than, for example, in blue [10]); see *fig. 9.*

---

[9]) W. T. Wintringham, Color television and colorimetry, Proc. Inst. Radio Engrs. **39**, 1135-1172, 1951.

Since yellow colours occur much more frequently than greenish blue, some improvement can be obtained by giving the green-fluorescent sulphide a somewhat higher cadmium-sulphide content, which shifts its colour point slightly towards yellow (in fig. 8 from G_s along the dot-dash line to the right).

Colour errors on a three-sulphide screen with an N.T.S.C. signal

A colour television system based, as far as green is concerned, on the colour point of willemite gives rise to colour errors on a screen that contains a green-fluorescent sulphide. These errors have been extensively studied, both in theory and practice, by Jackson [11]). His conclusion is that the errors can be almost completely compensated by means of a linear matrix network: all that is necessary is to subtract a specific percentage of the "green" signal from the "red" and "blue" signals. Especially in the absence of such compensation, it is advisable to shift the colour point of the red slightly towards orange. This has the additional advantage of improving the efficiency.

A favourable choice appears to be a combination of sulphides having the following cadmium-sulphide contents:

Colour	Phosphor	Colour co-ordinates	
		x	y
red	(ZnCd)S with 84% CdS	0.655	0.345
green	(ZnCd)S with 32% CdS	0.255	0.575

Blue remains unchanged.

The colour errors appearing on a three-sulphide tube under N.T.S.C. operation should not in any case be exaggerated. On repeated occasions we have compared such tubes with others containing the conventional phosphors. The same signal was fed to all monitors, and no correction was applied for differing phosphors. Although it could clearly be seen during the display of the primary colours that the green was less saturated on the three-sulphide tubes and that the red was more orange than on the normal tubes, hardly any difference was noticeable in a colour picture.

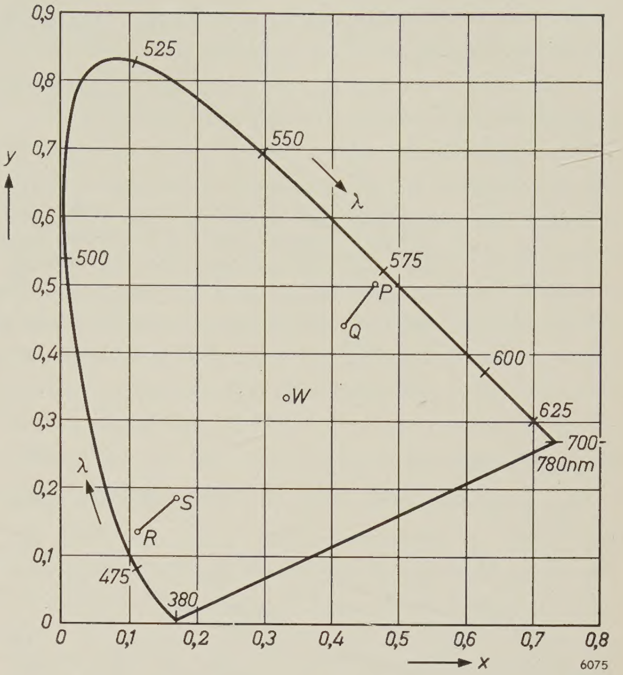


Fig. 9. Chromaticity diagram with two points in the yellow, P and Q, which are just as far apart as the points R and S in the blue. Between the yellow colours corresponding to P and Q the eye sees less difference than between the blue colours corresponding to R and S [10]).

To produce white on the new tubes, almost the same current ratios $R : G : B$ are needed as in existing tubes.

Summary. In direct-viewing tubes for colour television, use has hitherto been made of a phosphate for the red-fluorescent phosphor and of a silicate for the green-fluorescent phosphor. The sulphide (ZnCd)S-Ag, which fluoresces red or green depending on the CdS content, has substantially better properties (lower persistence, higher efficiency). Applied by the conventional method, however, the red-fluorescent sulphide does not adhere well to the tube face.

A new method giving good adhesion is described, in which a suspension containing the red-fluorescent sulphide is hardened by ultraviolet irradiation through the glass instead of from the gun side. The other phosphor patterns are coated with a dye to prevent the red phosphor from sticking to them. The result is a screen on which moving images are much sharper and which has 40 to 50% higher luminance.

---

[10]) D. L. MacAdam, Proc. Inst. Radio Engrs. **39**, 479 (fig. 12), 1951.

[11]) Unpublished investigation by R. N. Jackson of Mullard Research Laboratory, Salfords (England).

CIRCUITS FOR DIFFERENCE AMPLIFIERS, I.

by G. KLEIN *) and J. J. ZAALBERG van ZELST *).

621.375:621.317.725.083.6

Pursuant to an article recently published in this journal, which dealt with difference amplifiers in general, part I of the present article gives details of circuits designed to achieve the very high rejection factors that are frequently required. Part II, to be published in the next number, will offer hints on the efficient use of difference amplifiers. Some cases will also be described where difference amplifiers can be used with advantage even where the object is not to amplify a potential difference between two points.

Introduction

The behaviour of a difference amplifier is largely governed, as discussed in an earlier article ¹), by the characteristics of the first stage. In dealing now with various circuits used for difference amplifiers, we shall therefore be primarily concerned with single-stage amplifiers. The problems that arise when stages are added, some of which will also be touched on here, are as a rule not difficult to solve.

It was shown in the previous article that difference amplifiers are mainly used for amplifying low-frequency or DC signals. Our considerations will therefore be confined to such amplifiers, and we shall disregard the problems encountered at higher frequencies, connected for example with the capacitances of valves and other components.

The nature of the applications of difference amplifiers makes it necessary to stipulate for the most important characteristics — the rejection factor and the discrimination factor — a lower limit which differs from case to case. *The circuits that we shall deal with have been designed so as to be able to guarantee these minimum values without readjustment of the amplifier, even when the parameters of the valves and other components, which always show some mutual disparity, have the maximum deviation from the normal value and these deviations are all operative in the same (adverse) direction.*

As explained at some length in the above-mentioned article ¹), the problem of the difference amplifier consists in amplifying the voltage between two points which may both have a much higher potential with respect to earth. An obvious method of amplifying the potential difference between the two points would be to connect them with the input terminal and the "earth terminal" of a normal single-ended amplifier. The latter terminal is then not connected to earth; it can be said that electrically the whole amplifier "floats". Without going into details, it may be noted that amplifiers made electrically floating in this way are usually complicated and unwieldy in construction, and moreover usually have to be screened against interfering induction voltages. The principle described in this article, of *a balanced amplifier with a very high common cathode resistance*, offers in almost every case a much simpler solution.

Circuit with common cathode resistance

It was shown in the article quoted ¹) that a balanced amplifier, consisting of two independently operating sections, cannot be used as a difference amplifier because its discrimination factor is equal to unity, whereas for most purposes a value of at least 100 is required. In principle, a circuit can be used where the two valves of a balanced amplifier are given a common cathode resistance without decoupling (*fig. 1*). As we shall see, this resistance, R_k, must be very much higher than that needed for giving the grid the normal negative bias. For

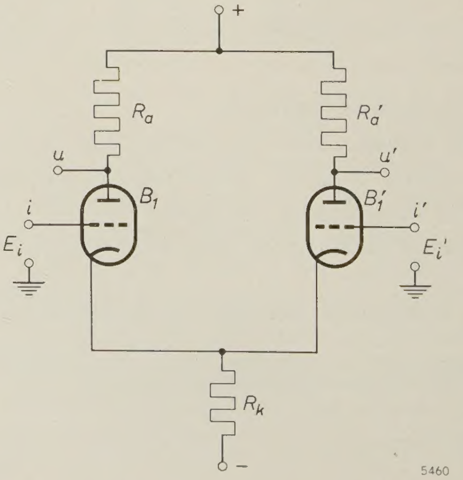


Fig. 1. Difference amplifier with a resistance R_k incorporated in the cathode lead common to both valves. Input terminals *i* and *i'*, output terminals *u* and *u'*.

---

*) Philips Research Laboratories, Eindhoven.

¹) G. Klein and J. J. Zaalberg van Zelst, General considerations on difference amplifiers, Philips tech. Rev. **22**, 345-351, 1960/61 (No. 11).

this reason, R_k is not usually connected to earth but to a voltage source which supplies a negative voltage with respect to earth. In this way the grids of valves biased to their normal operating point can be more or less at earth potential, which simplifies connection to the points whose voltages are to be measured, especially in the case of DC amplifiers.

If both valves are perfectly identical, and identically biased, there will be no change in the current through R_k when two equal but opposite small changes 2) are made to the two grid voltages E_i and E_i'. The resistance R_k then has no effect on the amplification of the two valves. If, however, the changes in E_i and E_i' have the same sign, the two valves function as if they were connected in parallel. The presence of R_k now gives rise to negative feedback. As a result the amplification of in-phase signals (see 1)) is smaller than that of anti-phase signals, and the discrimination factor F is therefore greater than unity. Assuming perfect symmetry, the value of F is easy to calculate. From the familiar expression for the gain of a triode the anti-phase gain is found to be

$$A = \mu \frac{R_a}{R_i + R_a}, \quad . \; . \; . \; . \quad (1)$$

where μ is the amplification factor and R_i the internal resistance of the valves, and R_a is the anode resistance. The in-phase gain C, where negative feedback due to R_k occurs, is given by

$$C = \mu \frac{R_a}{R_i + R_a + 2(1+\mu)R_k}. \quad . \; . \quad (2)$$

The discrimination factor is thus

$$F = \frac{A}{C} = 1 + 2\frac{(1+\mu)R_k}{R_i + R_a}. \quad . \; . \; . \quad (3)$$

In the circuits we shall be dealing with, the value of R_i is always large compared with R_a, and $2\mu R_k$ large compared with R_i. Since moreover the amplification factor μ is large compared with unity for normal valves, we can write F with negligible error as

$$F = 2SR_k, \quad . \; . \; . \; . \; . \; . \quad (4)$$

where S is the transconductance of both valves. Where the transconductance of the valves is not exactly the same, the discrepancy can be taken into account by inserting in eq. (4) the average value of S for both valves. Small discrepancies in the values of R_i and R_a are found to have an entirely negligible influence on F.

To compute the rejection factor H we must start from differences between the parameters, which cause a degree of asymmetry in the amplifier. If the amplifier were perfectly symmetrical, the value of H would of course be infinite. To find the rejection factor of a circuit as in fig. 1, we should therefore assume that the two halves of the amplifier differ in the transconductance and amplification factor of the valves and also in the resistance in the anode leads. The result of this calculation 3), provided the differences are not excessive, can be presented to a good approximation by the formula:

$$H = \frac{4}{\left(\frac{\Delta S}{S} + \frac{\Delta R_a}{R_a}\right)\frac{1}{SR_k} + \frac{1}{\mu}\left(\frac{\Delta\mu}{\mu}\right)\left(2 + \frac{R_a}{R_k}\right)}, \quad . \quad (5)$$

where ΔS, $\Delta\mu$ and ΔR_a are the differences in the transconductance, amplification factor and anode resistance of the valves respectively. Distinguishing the relevant values for the one valve from those for the other by a prime, we may write $\Delta S = S' - S$, $\Delta\mu = \mu' - \mu$ and $\Delta R_a = R_a' - R_a$.

From equation (5) we see that the smallest value of H occurs when $\Delta S/S$, $\Delta\mu/\mu$ and $\Delta R_a/R_a$ have the same sign (i.e. when S, μ and R_a of one valve are greater than those of the other valve), and moreover have the maximum values that can be expected from the normal tolerances of valves and resistors. The equation is greatly simplified if we assume that the maximum relative difference of the above three quantities have the same value, denoted by δ. (In practice, δ may for example be 0.1.) Inserting this in eq. (5) we find the minimum rejection factor that can occur for a given magnitude of δ:

$$H_{\min} = \frac{4}{\delta\left\{\frac{2}{SR_k} + \frac{1}{\mu}\left(2 + \frac{R_a}{R_k}\right)\right\}}. \quad . \; . \quad (6)$$

From this expression we see that $H_{\min}$ is greater the larger the value of R_k in the common cathode lead. If the value of R_k is so high that R_a/R_k is negligible compared to 2, we can simplify eq. (6) to:

$$H_{\min} = \frac{2}{\delta\left(\frac{1}{SR_k} + \frac{1}{\mu}\right)}. \quad . \; . \; . \quad (7)$$

Equations (6) and (7) show that to obtain a high value of $H_{\min}$ it is necessary, though not sufficient,

---

2) The changes must be small enough to prevent the curvature of the characteristics from playing any part.

3) See G. Klein, Rejection factor of difference amplifiers, Philips Res. Repts **10**, 241-259, 1955.

to have a large R_k. If SR_k is of the same order of magnitude as μ, then R_k must be very considerably increased to achieve a relatively slight increase in H_{min}. For a particular value of H_{min}, both SR_k and μ must have a specific minimum value which is larger the larger is H_{min}. If SR_k is small compared to μ, the minimum rejection factor is given by:

$$H_{min} = 2SR_k/\delta\,, \quad \ldots\ldots \quad (8)$$

or, putting $\delta = 0.1$:

$$H_{min} = 20\,SR_k\,. \quad \ldots\ldots \quad (9)$$

From the expressions (4) and (9) we see that in this case F is roughly equal to 0.1 H_{min}.

It may also happen that SR_k is large compared to μ, in which case H_{min} is approximately given by:

$$H_{min} = 2\mu/\delta\,, \quad \ldots\ldots \quad (10)$$

or, with $\delta = 0.1$:

$$H_{min} = 20\,\mu\,. \quad \ldots\ldots \quad (11)$$

In order to guarantee a rejection factor of, for example, 10 000 one must — assuming that R_k is sufficiently large — use valves having an amplification factor of at least 500. Where SR_k is not large compared to μ, the latter value must be even higher. For instance, H_{min} can also be made equal to 10 000 when both SR_k and μ are equal to 1000. Given a transconductance of 1 mA/V [4]) for both valves, R_k must then be equal to 1 MΩ.

In certain cases where an even higher rejection factor is required, a much higher value of R_k is needed, e.g. 10 MΩ. The use of such a resistor of normal construction in the common cathode lead is invariably inadvisable, in view of the abnormally high negative biasing voltage then called for. It is a fortunate circumstance, however, that the only requirement imposed on R_k is that the quotient of the voltage change and the resultant current change should be high; in other words, a high *differential* resistance is wanted. The DC resistance (voltage divided by current) may permissibly be very much lower and indeed should be so, having regard to our remarks on the supply voltage needed.

In one of the following sections we shall consider circuits which in fact combine a very high differential resistance with a low DC resistance. First, however, we shall examine in more detail various problems arising from the necessity of using valves having a high amplification factor.

Difference amplifiers using pentodes

Most pentodes have a much higher amplification factor than conventional triodes. Where valves with a high amplification factor have to be used in a difference amplifier, our thoughts first turn therefore to the use of pentodes. The circuit diagram of a difference amplifier with pentodes is shown in *fig. 2*. (For the present it is sufficient to show a normal resistance R_k in the common cathode lead.)

The very high amplification factor of a pentode is only used to full advantage when, given a variable control-grid voltage, the potential difference between screen grid and cathode is kept constant and when changes in screen-grid current flow only partly or not at all through R_k. If the difference amplifier is intended solely for alternating voltages, this can be achieved by connecting a capacitor C between the screen grids and the cathodes in the usual way (see fig. 2). Where the amplifier is also required to deal with DC signals, a voltage-stabilizing valve can be used instead of a capacitor (see *fig. 3*). Since the operating voltage of such a valve is very little dependent on the current, a practically constant potential is maintained in this way between the screen grids and the cathode, and changes in screen-grid current do not flow through R_k to any significant extent.

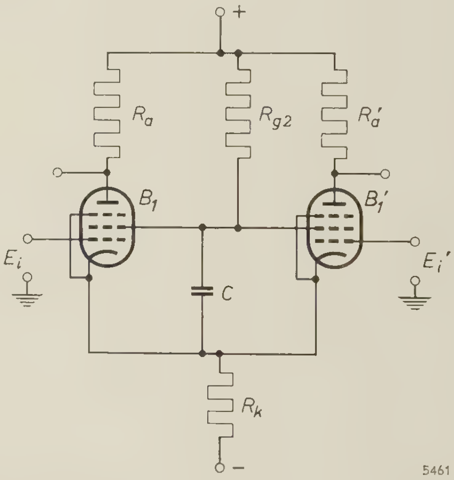


Fig. 2. Difference amplifier for AC signals, with two pentodes.

The above two conditions are never entirely fulfilled, for which reason H_{min} is smaller than follows from equation (7). The denominator of the complete equation in fact contains two extra terms, the first of which is due to the fact that the control-grid voltage and the screen-grid voltage do not necessarily influence the anode current in the same ratio in both valves. The second term is due to the possible spread in the distribution of cathode-current changes over anode and screen grid.

---

[4]) The valves are usually biased to obtain a low anode current, and hence a low mutual conductance, so that at the required value of R_k a lower negative supply voltage can be used. When the anode current is raised (and R_k correspondingly reduced) the mutual conductance increases relatively less, so that in spite of the higher S the product SR_k is no greater than when S is smaller.

A serious difficulty entailed by the use of pentodes arises from the flow of direct current to the screen grid. To obtain the required screen-grid potential the resistance R_{g2} must not exceed a specific value. (In fig. 3 the current of the voltage-stabilizing valve B_2 also flows through R_{g2}. In this case, therefore, the resistance in question must be even smaller than when a capacitor is used.) Since a constant potential difference is maintained between the screen grids and the cathodes, and also between the points marked + and —, the resistor R_{g2} may be regarded as in parallel with R_k as far as voltage and current changes are concerned. The differential resistance between cathode and earth is consequently reduced, and this causes, as we have shown above, a smaller value for the guaranteed rejection factor and discrimination factor. Methods that largely overcome this drawback will be discussed in the following sections.

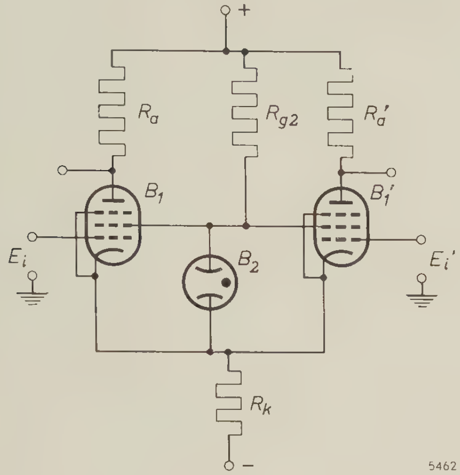


Fig. 3. Difference amplifier for DC signals, with two pentodes. A voltage-stabilizing valve B_2 is inserted between the screen grids and the cathodes.

The use of cascodes

An amplifier stage giving an amplification factor much higher than that of a triode can also be obtained by connecting two triodes in such a way as to produce a "cascode" circuit. The principle of such an arrangement is shown in *fig. 4*. The cathode of triode B_2 is connected to the anode of B_1. Provided the biasing voltages are so chosen that the valves operate in the normal part of their characteristics, a cascode exhibits properties closely resembling those of a pentode. The grid of triode B_2 functions in the circuit very much like the screen grid of a pentode. An important difference, however, is that screen-grid current always flows in a pentode, whereas in the "upper" valve in a cascode circuit no more than the usual, very low, grid current flows. This accounts for a marked advantage gained by using cascodes in difference amplifiers, to which we shall presently return.

A simple calculation shows that the transconductance of a cascode is practically identical with that of the "lower" valve:

$$S_{\text{casc}} = S_1 . \qquad \ldots\ldots \quad (12)$$

The amplification factor μ_{casc} of the whole circuit is given by

$$\mu_{\text{casc}} = \mu_1(\mu_2 + 1) , \qquad \ldots\ldots \quad (13)$$

where μ_1 and μ_2 are the amplification factors of the "lower" and "upper" valves, respectively. Since the amplification factors of normal valves are always very much larger than unity, we can write (13) to a very good approximation as

$$\mu_{\text{casc}} = \mu_1 \mu_2 . \qquad \ldots\ldots \quad (14)$$

The amplification factor of a cascode is thus high compared with that of each of the two valves of which it is composed.

A higher amplification factor, if required, can be obtained by using a cascode with more than two triodes. *Fig. 5* shows an example using three triodes. The gain of the cascode in this case is

$$\mu_{\text{casc}} = \mu_1(\mu_2 + 1)(\mu_3 + 1),$$

or, to a close approximation,

$$\mu_{\text{casc}} = \mu_1 \mu_2 \mu_3 .$$

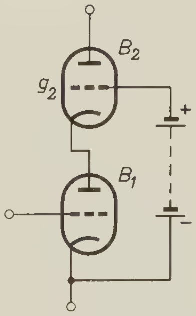


Fig. 4

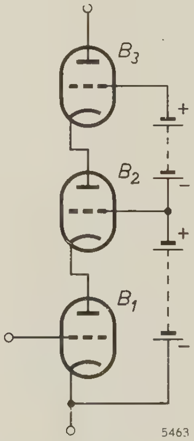


Fig. 5

Fig. 4. Cascode consisting of two triodes.

Fig. 5. Cascode consisting of three triodes.

To obtain an amplification factor as in (13) with a circuit as shown in fig. 4, the grid g_2 of B_2 must have a constant potential with respect to the cathode of B_1. This is represented in the figure, for simplicity, as being produced by a battery. As a rule, of course, the bias for g_2 will be derived from the anode-voltage source by means of a voltage divider. *Fig. 6* shows a voltage divider for this purpose in a difference amplifier consisting of two cascodes. Because of the extremely low grid current flowing in the valves, the resistance R_1 can be given a very high value, e.g. a few megohms. To maintain the required constant potential between the grids of B_2 and B_2' and the cathodes of B_1 and B_1' a capacitor C is introduced, the effect of which is to transfer practically all sufficiently rapid voltage variations from the lower cathodes to the upper grids. For alternating voltages, then, resistor R_1 can be regarded as being in parallel with R_k, so that the differential resistance in the common cathode lead is again smaller than R_k. In view of the fact, however, that R_1 can be given a very high value, this drawback is less serious here than with the screen grids of pentodes.

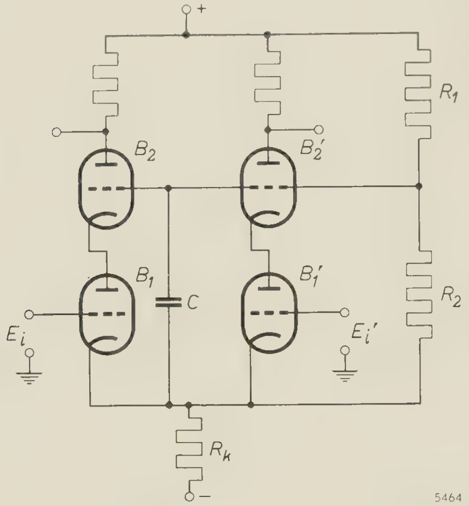


Fig. 6. Difference amplifier for AC signals, with two cascodes.

Cascodes are not so superior to pentodes when the difference amplifier is to be used for amplifying DC potential differences in the same way as the pentode circuit in fig. 3. The capacitor C will then be replaced by a voltage-stabilizing valve, and because of the direct current flowing in this valve the value of R_1 has to be made very much smaller. As in pentode circuits, arrangements can again be made that largely overcome this drawback. We shall touch on this under another heading.

It should be noted that the gain in $H_{\min}$ obtained by using cascodes instead of pentodes is less than would follow from equations (7) and (14). This is because the maximum spread in the amplification factor of a cascode is greater than that shown by the amplification factors for each valve individually. If μ_1 and μ_2 may each have a relative deviation δ from the nominal value, then the maximum relative deviation of μ_{casc} is equal to 2δ. When cascodes are used, therefore, the minimum guaranteed value of H is given not by equation (7) but by

$$H_{\min} = \frac{2}{\delta\left(\dfrac{1}{S_1 R_k} + \dfrac{2}{\mu_{\mathrm{casc}}}\right)} . \quad \quad (15)$$

The term accounting for the influence of the amplification factor on $H_{\min}$ is thus only reduced by half the ratio between the amplification factors when we change over from triodes to cascodes.

It should be remarked that equation (15), like equation (7) in the case of pentodes, is only valid on the assumption that there are no voltage variations between the grids of the upper valves and the cathodes of the lower ones. In order to allow for the fact that this is never the case, and assuming that the voltage variations on the grids are k times those on the cathodes ($k < 1$), the value for μ_{casc} to be inserted in (15) should not be calculated from (13) or (14) but from:

$$\mu_{\mathrm{casc}} = \frac{\mu_1(\mu_2 + 1)}{(1-k)\mu_2 + 1} .$$

Since hardly any grid current flows in a cascode, $H_{\min}$ is not reduced by any spread in the distribution of the cathode current, as it is in the case of pentodes.

It may sometimes be regarded as a disadvantage of cascodes that the power-supply circuit has to deliver a higher voltage than for pentodes. This is a particular objection in multi-stage DC amplifiers with direct coupling between the stages (see part II of this article). It is even more of a drawback, of course, when cascodes consisting of more than two valves are used.

Circuits for obtaining a high differential resistance in the cathode lead

We have seen from a numerical example that the use of a normal resistance in the common cathode lead of the valves in a difference amplifier seldom deserves consideration, but that a component or circuit is required for this purpose whose DC resistance is low and whose differential resistance is very high. This requirement can largely be met by incorporating a pentode in the cathode lead (*fig.* 7). If the control

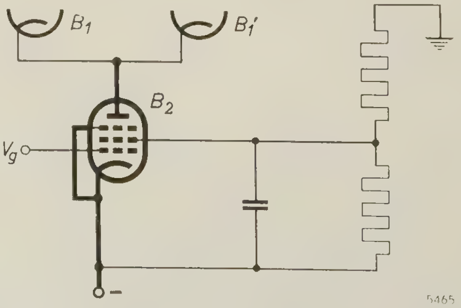


Fig. 7. Use of a pentode B_2 for producing a high differential resistance in the common cathode lead of the tubes B_1 and B_1'.

grid and screen grid are kept at the stipulated constant potentials with respect to the cathode, the differential resistance of a pentode is equal to its internal resistance, which may be more than 1 MΩ.

Another method of producing a high differential resistance in the common cathode lead is illustrated in *fig. 8*. The cathode lead of triode B_2 contains a resistance R_{k2}. If the grid of B_2 has a constant potential, the differential resistance R_d of the branch represented by thick lines in fig. 8 is given by

$$R_d = R_{i2} + (1 + \mu_2)R_{k2}, \quad . \quad . \quad . \quad (16)$$

where R_{i2} is the internal resistance and μ_2 the amplification factor of triode B_2. From eq. (16) we see that R_d is greater than $\mu_2 R_{k2}$. If, for example, the triode used has an amplification factor of 50 and R_{k2} is equal to 0.1 MΩ, then R_d will be greater than 5 MΩ.

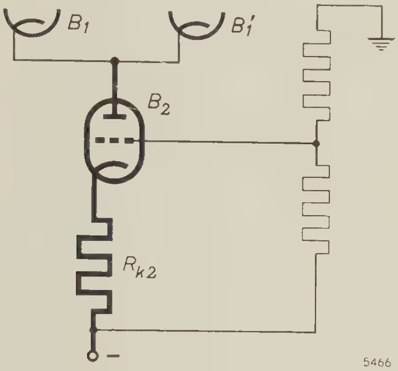


Fig. 8. Use of a triode B_2 with cathode resistance R_{k2} for producing a high differential resistance in the common cathode lead of the tubes B_1 and B_1'.

A pentode has a much higher amplification factor than a triode, and therefore R_d can be given a higher value if the triode B_2 in fig. 8 is changed for a pentode. We should bear in mind, however, that equation (16) is only valid for a pentode provided the changes in the anode current are equal to the changes in the current through the resistance in the cathode lead. Since there is always some screen-grid current flowing in a pentode, the only way to fulfil this condition is to insert between the screen grid and cathode either a capacitor of sufficiently high capacitance or a gas-discharge tube. In many cases it is then simpler to use a cascode arrangement, where there is no grid current and therefore no need for the above measure.

Fig. 9 shows a cascode, formed from two triodes B_2 and B_3, incorporated in the common cathode lead of valves B_1 and B_1'. The cathode lead of B_2 contains the resistor R_{k2}, and the grid voltages are kept constant by a voltage divider R_1-R_2-R_3.

We have seen that the amplification factor of a cascode is greater than the product of the amplification factors of the two valves. For this reason the differential resistance of the branch represented by thick lines in fig. 9 is greater than $\mu_2\mu_3R_{k2}$ (μ_2 and μ_3 being the amplification factors of triodes B_2 and B_3). Here too it is possible to use cascodes built up from more than two triodes, so that there is practically no limit to the value of the differential resistance which can be obtained in this way.

As explained above, the differential resistance in the common cathode lead of valves B_1 and B_1' does not consist solely of the thickly drawn branch in figures 7, 8 and 9, but also of a resistance in parallel with this branch and represented by R_{g2} in figures 2 and 3 and by R_1 in fig. 6. Since R_{g2} is much smaller than the required differential resistance (in DC amplifiers this also applies to R_1, as we have seen above), this is a severe obstacle to a high rejection factor. Two circuits that largely overcome this obstacle are represented in *figs 10* and *11*.

In fig. 10 the required constant potential difference between the cathodes and screen grids of B_1 and B_1' is not obtained by interposing a capacitor or voltage-stabilizing valve, but by means of a cathode follower B_3 and a coupling capacitor C. (Where the difference amplifier is also to amplify DC potential differences, a voltage-stabilizing valve should be substituted for the capacitor C.) If B_3 were an ideal cathode follower and the reactance of the capacitor C were negligible compared with the resistance R_{g2}, the voltage on the screen grids would completely "follow" the voltage variations on the cathodes. In reality, of course, this is never so. For that to be possible the differential resistance between the

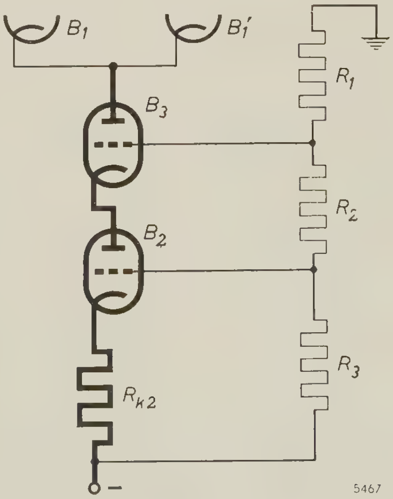


Fig. 9. Cascode with cathode resistance R_{k2} used for producing a high differential resistance in the common cathode lead of the tubes B_1 and B_1'.

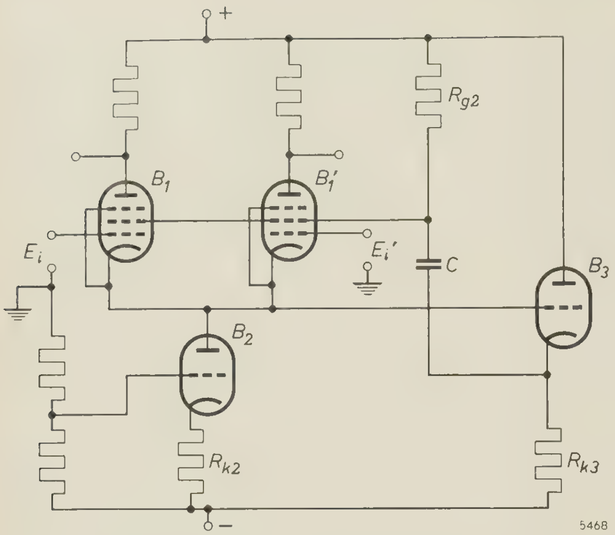


Fig. 10. Difference amplifier with pentodes B_1 and B_1'. The common cathode lead is given a high differential resistance by means of the triode B_2 and the resistor R_{k2}. The cathode follower B_3 enables the screen grids to follow the voltage fluctuations on the cathodes of B_1 and B_1'.

cathode of B_3 and earth would have to be infinitely high, and B_3 itself would have to have an infinitely large amplification factor. There will still, therefore, be slight voltage variations between the screen grids and the cathodes of B_1 and B_1', but owing to the fact that R_{g2} in fig. 10 has no influence on the common cathode resistance of these valves the cathode follower B_3 does in fact considerably improve the rejection factor and the discrimination factor.

The situation is even more improved if the cathode follower B_3 in fig. 10 is replaced by an amplifier whose gain approaches still closer to unity. In this way the influence of R_{g2} on the rejection and discrimination factors can be almost entirely eliminated.

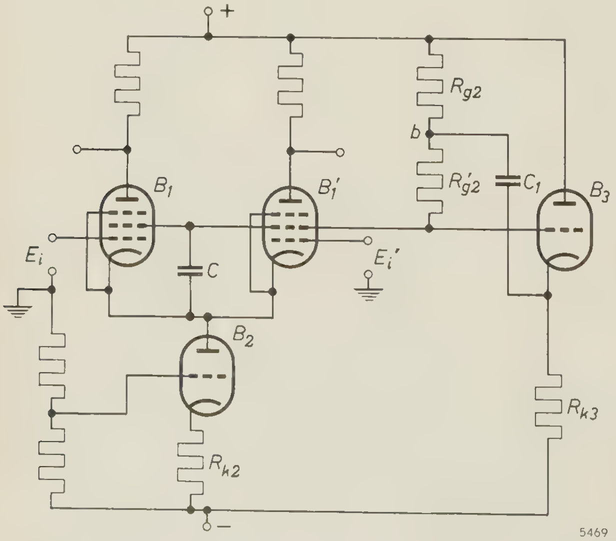


Fig. 11. Difference amplifier with pentodes B_1 and B_1'. The common cathode lead is given a high differential resistance by means of the triode B_2 and the resistor R_{k2}. The cathode follower B_3 raises the differential resistance between the screen grids and the anode-voltage source to a very high value.

In fig. 11 a capacitor C is again shown incorporated between the cathodes and screen grids of pentodes B_1 and B_1'. Given a high enough value of C, the voltage on the screen grids will almost completely follow the voltage variations on the cathodes. In that respect, then, this circuit has an advantage over that in fig. 10. The disadvantage, that the DC resistance of the screen grids reduces the differential resistance in the common cathode lead of B_1 and B_1', is largely overcome by the fact that the screen grids are here connected to the anode-voltage source by a circuit which has a very high differential resistance but a much smaller DC resistance. This circuit consists of the two resistors R_{g2} and R_{g2}' and a cathode follower B_3, which, via capacitor C_1, transmits the voltage variations on the screen grids almost entirely to point *b*. As a result the differential resistance between the screen grids and the anode-voltage source is much greater than $R_{g2} + R_{g2}'$, which can mean a considerable increase in the rejection and discrimination factors.

In this circuit too (fig. 11) the cathode follower B_3 can be changed for an amplifier whose gain is closer to unity. Use can also be made of circuits combining the principles of figs 10 and 11. Details of these circuits, however, are beyond the scope of this article.

The methods illustrated in figs 10 and 11 for increasing the differential resistance in the common cathode lead can also be adopted, of course, when cascodes are used in the differential amplifier instead of pentodes. As mentioned above, when cascodes are used these methods will generally be needed only when the differential amplifier is required to amplify DC potentials, in which case voltage-stabilizing valves must be used instead of capacitors, and the grids of the "upper" valves (see fig. 6) are therefore connected via a relatively small resistance to the anode-voltage source.

Some results of measurements

It will be clear from the foregoing that we can choose from a wide variety of circuit arrangements in order to design a difference amplifier whose rejection and discrimination factors can be guaranteed to have very high values. From the numerous possibilities, we have chosen three examples with a view to comparing the measured rejection factors with the minimum values calculated for these circuits.

Fig. 12 shows the circuit of a difference amplifier equipped with two E 80 F pentodes. The high differential resistance required in the common cathode lead is obtained with the triode B_2 and the resistor R_{k2}. By means of the cathode follower B_3 the voltage variations on the cathodes of B_1 and B_1' are transmitted to point *b*. The influence of the screen-grid resistance on the rejection factor is thus reduced here by a circuit which combines the principles illustrated in the figures 10 and 11. The two halves of a double triode ECC 81 are used for B_2 and B_3.

The minimum value of the rejection factor, assuming $\delta = 0.1$, is calculated to be 20 000 for this circuit. In *fig. 13* the calculated minimum is represented on a logarithmic scale together with the rejection factors measured on 25 arbitrary combinations of valves. The measurements were done at a frequency of 1 kc/s. The lowest value measured was 24 000.

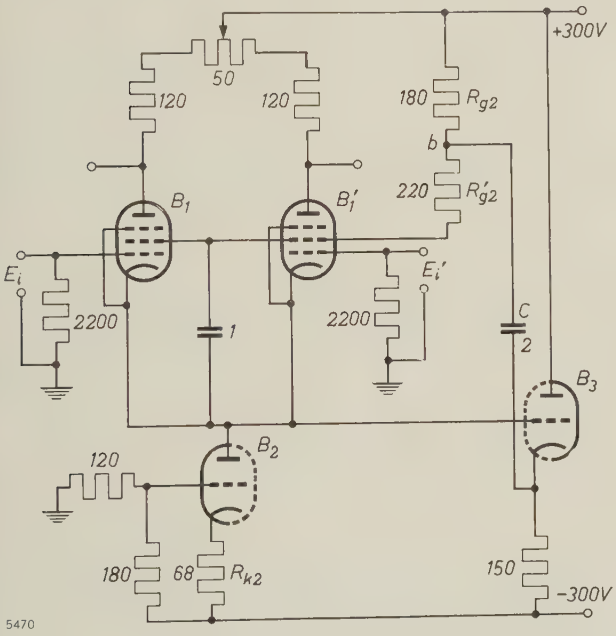


Fig. 12. Difference amplifier with two pentodes, type E 80 F. Triodes B_2 and B_3 are formed from the two halves of a double triode ECC 81. Resistances are given in kΩ, capacitances in μF. The 50-kΩ potentiometer is adjusted in such a way that in the steady state the anode voltages of the pentodes do not differ too much, giving both tubes roughly the desired operating point.

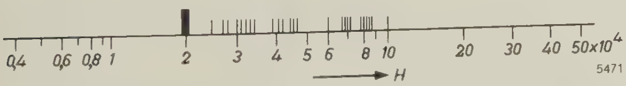


Fig. 13. Measured values of rejection factor *H* (thin marks) and calculated minimum value $H_{\min}$ (thick mark) for the circuit of fig. 12.

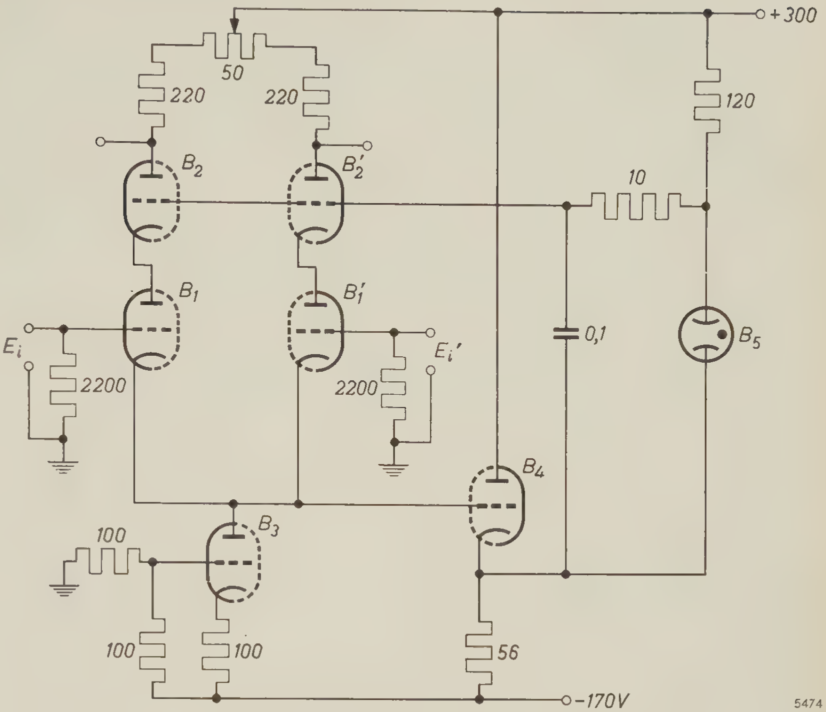


Fig. 14. Difference amplifier for DC signals, with two cascodes composed of two double triodes type E 80 CC. Triodes B_3 and B_4 are formed from the two halves of a double triode type ECC 81. B_5 is a voltage-stabilizing valve type 85 A 2. Resistances in kΩ, capacitances in μF.

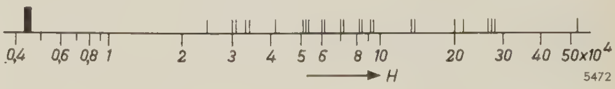


Fig. 15. Measured values of rejection factor *H* (thin marks) and calculated minimum value $H_{\min}$ (thick mark) for the circuit of fig. 14.

Fig. 14 gives the circuit diagram of a difference amplifier with two cascodes consisting of two double triodes, type E 80 CC. One half of a double triode ECC 81 is used for B_3 and the other half as a cathode follower, B_4, for transmitting the voltage variations on the cathodes of B_1 and B_1' to the grids of B_2 and B_2'. The transmission is effected by a voltage-stabilizing valve B_5 of type 85 A 2. Since the differential resistance of such a valve increases with increasing frequency, a capacitor of 0.1 μF is connected in parallel with B_5. The resistance of 10 kΩ between this capacitor and B_5 serves to correct instability effects likely to occur in a circuit of this kind.

Assuming that all quantities involved may show mutual deviations of 10%, we calculate a minimum value of 4500 for the rejection factor of this circuit. The values measured on 25 arbitrary combinations of valves are again shown on a logarithmic scale in *fig. 15*. The lowest value measured, 24 000, is much higher than the calculated minimum value.

The reason is that in fact the mutual disparity in the transconductance and amplification factor of a type E 80 CC double triode very seldom approaches 10%. Consequently there is a good chance that higher values of H will be found with circuits using this valve rather than other types whose S and μ are higher but show a greater mutual disparity.

The measurements whose results are represented in fig. 15 were carried out with direct voltage. If the difference amplifier is intended to handle alternating voltages only, the circuit can be considerably simplified. As already shown in fig. 6, the voltage for the grids of B_2 and B_2' can then be obtained from a voltage divider between the anode-voltage source and the cathodes of B_1 and B_1'. The voltage variations on the cathodes of the latter valves can be transmitted to the grids of B_2 and B_2' by means of a capacitor. *Fig. 16* represents the diagram of a circuit on which, with various combinations of valves, measurements were carried out at a frequency of 100 c/s. Two E 80 CC double triodes were used for B_1, B_1', B_2 and B_2', and half of a double triode ECC 81 was used for B_3. Again assuming 10% deviations in the values of all quantities involved, the minimum value calculated for the rejection factor was 6500. In *fig. 17* the measured values are again set out together with the calculated minimum value of H. The lowest value measured on 25 arbitrary combinations of valves was 24 000. Here, too, owing to the constancy of the E 80 CC double triode, this measured minimum is appreciably higher than the calculated value.

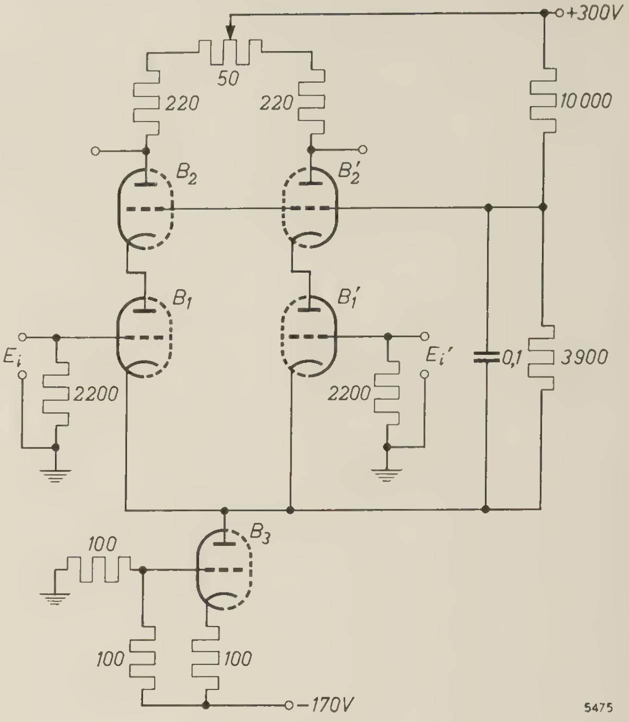


Fig. 16. Difference amplifier for AC signals, with two cascodes composed of two E 80 CC double triodes. B_3 is one half of an ECC 81 double triode. Resistances in kΩ, capacitances in μF.

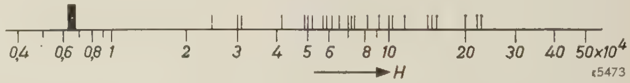


Fig. 17. Measured values of rejection factor H (thin marks) and calculated minimum value $H_{\min}$ (thick mark) for the circuit of fig. 16.

Part II of this article will deal with various problems that arise in the design of multi-stage difference amplifiers. Indications will also be given of the proper method of connecting the amplifier to points between which the potential difference is to be measured. Finally, some cases will be mentioned where difference amplifiers can also be used with advantage where the object is not to amplify the potential between two points.

Summary. An efficient single-stage difference amplifier can be produced by incorporating in the common cathode lead of the two valves of a balanced amplifier an element which has a very high differential resistance. The latter is necessary to obtain a high discrimination factor. In order to be able to guarantee a high rejection factor, even though the parameters of valves and other components show the maximum (adverse) deviation from their rated value, not only must the differential resistance in the cathode lead be high but the valves must also have a high amplification factor. The valves used may be either pentodes, or triodes in a cascode arrangement. The voltage on the screen grids of the pentodes or on the grids of the "upper" triodes in the cascodes must follow the voltage variations on the cathodes closely. Some circuits which achieve this result are described. The high differential resistance in the common cathode lead can be obtained with circuits using pentodes, triodes or cascodes. Measurements on three circuits are discussed, and the calculated minimum rejection factor is compared with the values measured on 25 arbitrary combinations of valves.

LOOP GAIN AND STABILITY OF SIMPLE CONTROL SYSTEMS

by M. van TOL *).

621-53.001

In the development of control theory during the last twenty years, two main lines may be distinguished. On the one hand refined mathematical methods have been introduced, such as the use of Laplace transforms, and on the other hand approximations have led to simple rules of thumb, which are of considerable practical value.

In the following article such a rule of thumb is given for finding the relationship between the time constants of a control system and the maximum loop gain which the stability of the system permits. This rule, which has not as far as we know been explicitly formulated elsewhere, makes it possible to show the relationship between the various methods employed to improve the performance of a control system.

The factor P by which a disturbance is reduced in an automatic control system depends on the transfer function $KG(j\omega)$ of the open control loop. In a previous article [1]) it was shown that P is equal to the absolute value of the complex quantity $1 + KG(j\omega)$ and therefore approaches $1 + K$ for very low frequencies. (We assume provisionally that $G(0)$ is equal to unity.) A step-function disturbance is therefore ultimately corrected by the system except for a portion $1/(1+K)$. Since this portion is smaller the larger the value of K, the aim is to make K as large as possible.

There is a limit, however, to the possible increase of K. Since the loop gain at any given frequency is proportional to K, a higher K results in a rise in gain at the frequency ω_c at which the phase shift φ is —180°. When this gain exceeds unity the system is no longer stable. In designing a control system — and the same applies to negative-feedback amplifiers and servo-mechanisms — it is therefore necessary to make $KG(0)$ high enough to reduce disturbances effectively, at the same time ensuring that $|KG(j\omega_c)| < 1$. The problem thus amounts to finding a suitable form for the function $G(j\omega)$. The measures taken to that end must also ensure that the response of the system is not impermissibly slow.

Control engineers have long used certain established methods of improving the performance of control systems. In this article we shall show that all these methods, between which there has hitherto seemed to be little or no connection, may be regarded as applications of one and the same principle. That principle is directly derived from a very simple formula which defines the relationship between the maximum permissible value of K and the two longest characteristic times τ that define the behaviour of the elements of the control loop [2]). Although this formula is not exact unless the control loop meets certain requirements, it can be shown to be of practical value in numerous cases where these requirements are not completely satisfied.

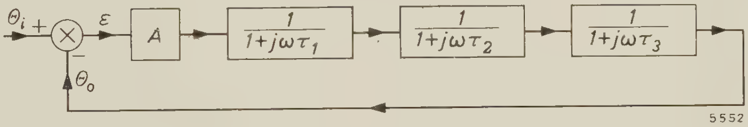

Θ_i
+
ε
A
$\frac{1}{1+j\omega\tau_1}$
$\frac{1}{1+j\omega\tau_2}$
$\frac{1}{1+j\omega\tau_3}$
—
Θ_o
5552

Fig. 1. Block diagram of a control system consisting of a linear amplifier, gain factor A, and elements having the transfer function $(1 + j\omega\tau_n)^{-1}$. In the usual terminology Θ_i is the desired value, Θ_o the instantaneous value of the controlled condition (output quantity) which is compared with Θ_i, and ε, the deviation, is their difference.

The basic formula

Let us consider a control loop that can be represented by the block diagram in *fig. 1*. Apart from the controlling unit, a linear amplifier with gain factor A, we may write for all blocks:

---

*) Research Laboratories, Eindhoven.

[1]) M. van Tol, Application of control theory to linear control systems, Philips tech. Rev. **23**, 109-118, 1961/62 (No. 4).

[2]) See also: M. van Tol, Stability and optimal loopgain of simple control circuits, Transactions of the 5th International Instruments and Measurements Conference, Stockholm 12-16 Sept. 1960, or M. van Tol, Control engineering **8**, 1961, in the press.

$$K_n = 1, \quad \dots\dots\dots \quad (1)$$

$$G_n = \frac{1}{1 + j\omega\tau_n}. \quad \dots\dots \quad (2)$$

The amplitude-frequency response characteristics of each of these blocks (which we shall here refer to as amplitude characteristics for short) when plotted logarithmically in a Bode diagram, approximate to two straight lines intersecting on the vertical $\omega = 1/\tau_n$. The first, which approximates to the low-frequency characteristic, is horizontal; the other has a slope of -1. The phase shift φ between output and input signal varies with increasing ω from 0 to $-90°$, and at $\omega = 1/\tau_n$ is exactly $-45°$ (*fig. 2*).

The slope of these two lines can be found by writing the differential quotient d log $|G|$/d log ω as a function of ω, and then ascertaining how this function behaves where $\omega \ll 1/\tau_n$ and where $\omega \gg 1/\tau_n$. Using the formula

$$|G| = 1/\sqrt{1 + \omega^2\tau_n^2},$$

we find

$$\frac{\mathrm{d}\log |G|}{\mathrm{d}\log \omega} = \frac{-\omega^2\tau_n^2}{1 + \omega^2\tau_n^2}.$$

Where ω is small the differential quotient is thus ~ 0; where ω is large it approaches -1. The line $|G| = 1/\omega\tau_n$ intersects the other at the frequency for which the value of the ordinate is unity. As can be seen, that frequency is $\omega = 1/\tau_n$.

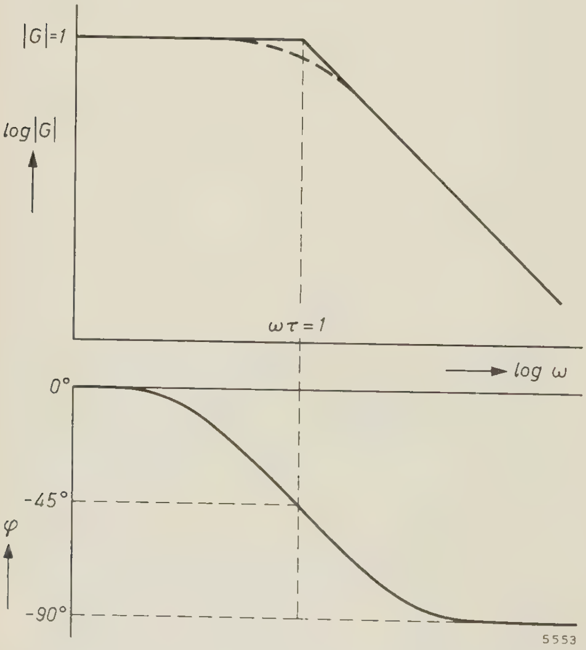


Fig. 2. Bode diagram of an element having the transfer function $G = (1 + j\omega\tau)^{-1}$. The amplitude-frequency response characteristic is approximated by two straight lines intersecting at the vertical $\omega = 1/\tau$. The phase shift φ is exactly $-45°$ at the frequency $1/\tau$ and has a maximum value of $-90°$.

The amplitude characteristic in logarithmic coordinates of the complete open loop can very simply be derived from that of the individual blocks, every multiplication being an addition on logarithmic paper. This characteristic, too, can thus be approximated by straight lines.

If the various time constants τ_n are found to differ considerably — e.g. by a factor of 5 or more — the phase-frequency response characteristic (or phase characteristic for short) can also be found to a good approximation by adding the phase characteristics of the individual blocks. The phase shift is then roughly $-45°$ for $\omega = 1/\tau_1$, about $-135°$ for $\omega = 1/\tau_2$, about $-225°$ for $\omega = 1/\tau_3$ and so on (*fig. 3*).

As mentioned above, for a stable system $|KG|$ should be less than unity at the frequency ω_c where φ is $-180°$. In practice, this limit should not be too closely approached; the closer the phase shift at the frequency for which $|KG| = 1$ approaches $-180°$, the poorer is the damping of the system and therefore the longer the oscillations last. For practical purposes a phase "margin" of 45° is found to be sufficient to guarantee stability. In other words, the loop gain drops below unity at the frequency where the phase shift is $-135°$. In the event of a step-function transient the overshoot Θ_s' is then no more than about 15% (see *fig. 4*). As can be seen from fig. 3, a phase shift of $-135°$ occurs at exactly $\omega = 1/\tau_2$. To meet the stability requirement, then,

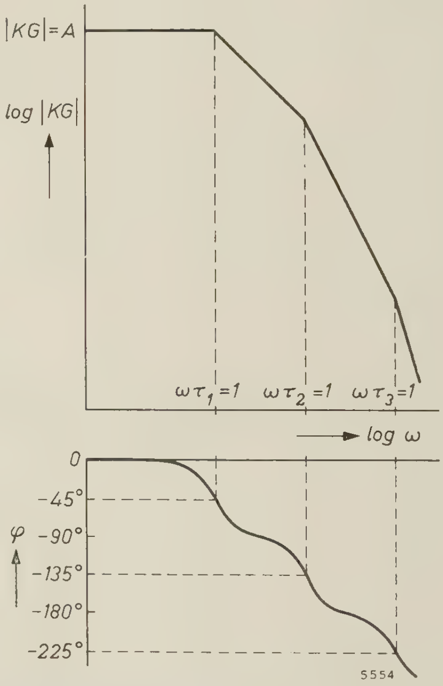


Fig. 3. Bode diagram (approximation) of an open control loop consisting of a linear amplifier (gain A) and various elements having the transfer function $(1 + j\omega\tau_n)^{-1}$, where $\tau_1 \gg \tau_2 \gg \tau_3$ etc.

A must not be greater than the value $A_{\max}$ which causes the response to reach unity at the second break in the amplitude characteristic (*fig.* 5).

As we have seen, the schematic amplitude characteristic of a single block has a slope of -1 for frequencies greater than $1/\tau_1$. The same therefore holds for the section PQ (fig. 5) of the amplitude characteristic of the entire (open) control loop. The triangle PQR is thus isosceles: $PR = QR$. Hence:

$$\log A_{\max} - \log 1 = \log 1/\tau_2 - \log 1/\tau_1, \qquad (3)$$

or

$$A_{\max} = \frac{\tau_1}{\tau_2}. \qquad (4)$$

Put into words: *to ensure stability, the open loop gain must not be greater than the quotient of the longest two time constants.*

The principle underlying nearly all methods that can be used to improve the performance of an automatic control system follows directly from equation (4). If the system is not sufficiently stable, then A is too large, or in other words the quotient τ_1/τ_2 is too small. The latter must then be increased until it is at least equal to A. If disturbances are not adequately suppressed, then A must be increased, but at the same time, to maintain stability, τ_1/τ_2 must also be raised so that it remains at least equal to A. *In both cases, then, the rule is to try to increase the quotient* τ_1/τ_2.

Although, as appears from the derivation of eq. (4), this rule applies only to control loops where all blocks have the transfer function $K_n/(1 + j\omega\tau_n)$, and where $\tau_1 \gg \tau_2 \gg \tau_3$ etc., it is nevertheless a useful starting point for many practical cases.

Applications

The methods commonly used to improve an unsatisfactorily operating control system are: to change one of the time constants (if applicable, an exceptionally simple method); to introduce an element giving a negative phase shift or one giving a positive phase shift; to introduce an integrating element or a differentiating element, or a combination of both.

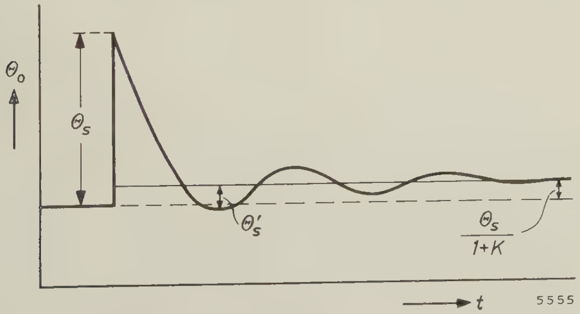


Fig. 4. With a "phase margin" of 45°, a step-function disturbance of magnitude Θ_S gives rise to an overshoot Θ_S' no greater than about 0.15 Θ_S.

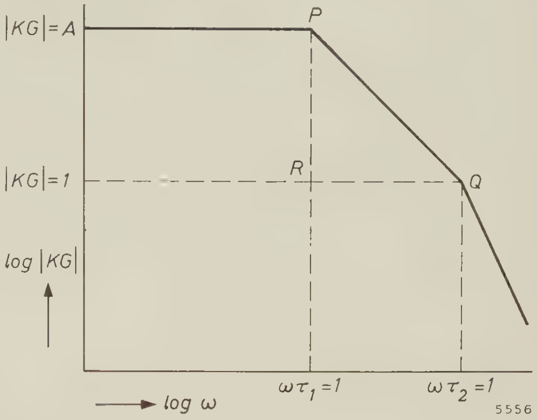


Fig. 5. With a phase margin of 45° the loop gain $|KG|$ is exactly equal to unity when $\omega = 1/\tau_2$, i.e. at the second break Q in the logarithmic amplitude-frequency response characteristic approximated by straight lines. The slope of the part between the first break (P) and the second is -1.

We shall now show that all these measures amount to increasing the quotient τ_1/τ_2. As far as the first method is concerned — changing one of the time constants — this is immediately evident: τ_1 is increased or τ_2 decreased (the latter is only possible, of course, provided τ_3 is sufficiently small.) The other methods require somewhat more detailed analysis.

Introduction of a phase-shifting element

The electrical or pneumatic circuits used for producing a negative phase shift for the above purpose — the electrical ones are known as phase-lag networks [3]) — have the transfer function

$$G' = \frac{1 + j\omega\tau_b}{1 + j\omega b\tau_b}, \qquad (5)$$

where $b > 1$. If we make τ_b equal to τ_1, the longest time constant of the control loop, the transfer function of the combination of the block with time constant τ_1 and the phase-lag network is:

$$G_1' = G' \times G_1 = \frac{1 + j\omega\tau_1}{1 + j\omega b\tau_1} \times \frac{1}{1 + j\omega\tau_1} = \frac{1}{1 + j\omega b\tau_1}. \qquad (6)$$

The unit with time constant τ_1 has thus as it were been replaced by a similar block having the time constant $b\tau_1$. After introducing such a network, we can therefore increase the gain by a factor b without reducing stability. If τ_1 is already very great it may not be possible to use this method, since it may not be possible to make a phase-shifting element with the even greater time constant $b\tau_1$.

---

[3]) A description of these networks will be found in books dealing with the theory of electrical circuits. See also e.g. G. S. Brown and D. P. Campbell, Principles of servomechanisms, Wiley, New York 1948, chapter 7; or R. A. Bruns and R. M. Saunders, Analysis of feedback control systems, McGraw-Hill, New York 1955, chapters 6 and 13.

Using a "phase-lead network" — also called a damping network because it generally increases the damping in the closed loop — the value of τ_2 is reduced in a corresponding manner. The transfer function of such an arrangement [3]) is:

$$K''G'' = \frac{1}{c}\,\frac{1 + j\omega\tau_c}{1 + j\omega\tau_c/c}, \quad . \quad . \quad . \quad . \quad (7)$$

where $c > 1$. Choosing $\tau_c = \tau_2$, the effect is as if the time constant τ_2 in the open loop had been changed to τ_2/c. The loop gain may therefore be increased by the factor c.

It is a condition here, as in the method where τ_2 itself is reduced, that τ_3 should not approach too closely to τ_2.

Since the introduction of the network in itself decreases the gain by a factor c, when this method is used the total extra gain must be c^2. The fact that the phase-lead network attenuates the signal may sometimes be a drawback; in some cases the attenuated signal may no longer be strong enough in relation to the input noise of the amplifier.

In practice, b and c in these two methods are given values of the order of 10.

The introduction of an integrating or differentiating element

The conventional method of introducing an integrating or differentiating element in a control loop is illustrated in *fig. 6*. The signal ε_o applied to block IV is not simply the A-times amplified deviation $\Theta_i - \Theta_o$, which we shall call ε_i, but contains a component originating from the integrating or differentiating element (block *III*). If block *III* is an integrating element, we speak of P.I. control (P = proportional, I = integral); if *III* is a differentiating element, we speak of P.D. control (D = derivative).

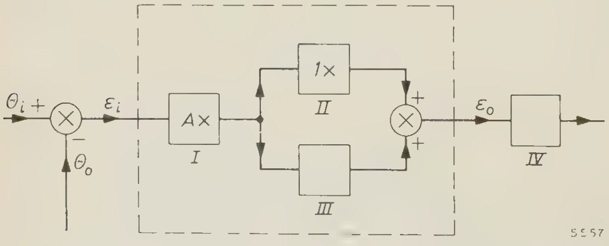


Fig. 6. Method of introducing an integrating or differentiating element (block *III*) in a control loop (P.I. control and P.D. control, respectively). *I* is a linear amplifier, gain factor *A*. *II* ditto, with gain factor equal to unity.

In the case of P.I. control we have:

$$\varepsilon_o = A\left\{\varepsilon_i + \frac{1}{\tau_i}\int_0^t \varepsilon_i \, dt\right\}. \quad . \quad . \quad (8)$$

For sinusoidal signals the transfer function of the whole assembly of units inside the dotted square is given by:

$$KG_{\mathrm{P.I.}} \equiv \frac{\varepsilon_o}{\varepsilon_i} = A\left(1 + \frac{1}{j\omega\tau_i}\right) = A\,\frac{1 + j\omega\tau_i}{j\omega\tau_i}. \quad (9)$$

Now the integrating action of practical networks is never perfect. Given a constant input signal, the output signal does not go on increasing indefinitely but finally reaches a constant value. In many cases the gain is not infinite at zero frequency, but has a finite value which we shall call B. The transfer function is then not $1/j\omega\tau_i$ but

$$\frac{1}{(1/B) + j\omega\tau_i}.$$

The transfer function $KG_{\mathrm{P.I.}}$ is therefore not given in such cases by eq. (9) but by

$$KG_{\mathrm{P.I.}} = A\left\{1 + \frac{1}{(1/B) + j\omega\tau_i}\right\} =$$
$$= AB^*\,\frac{1 + j\omega\tau_i^*}{1 + j\omega B^*\tau_i^*}, \quad (10)$$

where $B^* = B + 1$ and $\tau_i^* = B\tau_i/(B + 1)$.

It will be seen that this transfer function has the same form as that of a phase-lag network. If we choose τ_i such that $\tau_i^* = \tau_1$, then τ_1 in the loop is again apparently increased to $B^*\tau_i^*$, i.e. to $B\tau_i$. A difference is that the factor B can be much larger than the maximum feasible value of b.

The non-ideal behaviour of practical integrators can be briefly explained with the aid of two examples. The first is a capacitor charged by a current. The voltage cannot go on rising indefinitely for two reasons: *a*) the sheathing materials of the capacitor are not perfectly insulated from one another, and at very high tensions the leakage current is no longer negligible; *b*) no current source can continue to supply current under unlimitedly high inverse voltage. The same applies, *mutatis mutandis*, to pneumatic integrators. It makes no formal difference which of the two causes dominates in a given case. The second example is an electric motor that cannot go on turning indefinitely but must stop as soon as e.g. the sliding contact of the potentiometer which it operates reaches the end of the resistance. The cause of the non-ideal behaviour is of a different nature in both examples. In the first case we have an unwanted extra time constant, but the element remains linear; in the second example non-linearity occurs.

In both cases ideal operation is approached only when the output signal is not unduly large. In arriving at equation (10) we envisaged an integrator of the type of the first example.

In the case of P.D. control, the transfer function $KG_{\mathrm{P.D.}}$ is found from:

$$\varepsilon_{\mathrm{o}} = A\left(\varepsilon_{\mathrm{i}} + \tau_{\mathrm{d}}\frac{\mathrm{d}\varepsilon_{\mathrm{i}}}{\mathrm{d}t}\right). \quad . \quad . \quad . \quad (11)$$

For sinusoidal signals this gives:

$$KG_{\mathrm{P.D.}} \equiv \frac{\varepsilon_{\mathrm{o}}}{\varepsilon_{\mathrm{i}}} = A(1 + \mathrm{j}\omega\tau_{\mathrm{d}}). \quad . \quad . \quad (12)$$

Here again, the actual situation differs somewhat from the theoretical, for the differentiating action ceases when the frequency exceeds a particular value. The transfer function in practice is therefore given by:

$$KG_{\mathrm{P.D.}} = A\,\frac{1 + \mathrm{j}\omega\tau_{\mathrm{d}}}{1 + \mathrm{j}\omega\tau_{\mathrm{d}}/C}, \quad . \quad . \quad . \quad (13)$$

where $C \gg 1$.

Just as with the phase-lead network, the apparent effect is a reduction of τ_2. Again the difference is that C can be very much greater than the factor c. Both B and C are usually of the order of 100.

As mentioned, a combination of two or more of the methods discussed above can be used. A familiar example is a P.I.D. controller, a device which has both an integrating and a differentiating element in parallel with the linearly amplifying element. In this case we should choose $\tau_{\mathrm{i}} = \tau_1$ and $\tau_{\mathrm{d}} = \tau_2$. This too constitutes a convenient rule of thumb for adjusting the controller in many of the situations encountered in practice.

A simple rule can also be given for the effect which the measures discussed have on the speed with which the control system reaches the new steady state after a step-function disturbance. The rule is: *if the gain A is kept equal to τ_1/τ_2, an increase in τ_1 at constant τ_2 does not change the speed of response, whereas a decrease in τ_2 at constant τ_1 makes the response faster.* The truth of this can quickly be verified. In broad terms, the response time of low-pass filters — to which category controlled systems belong — is equal to the reciprocal of the highest frequency f_0 passed through [4]). Now if A is τ_1/τ_2 the highest transmitted angular frequency ω_0 is given by $\omega_0\tau_2 = 1$. This applies not only to the open loop (cf. fig. 5) but equally to the closed loop, whose amplitude characteristic is given by the formula:

$$\frac{\Theta_{\mathrm{o}}}{\Theta_{\mathrm{i}}} = \frac{KG}{1 + KG}.$$

The response time is therefore proportional to τ_2 and is not affected by variations in τ_1 — at any rate as long as τ_1 is several times larger than τ_2.

Summarizing, it may be said that the basic formula $A_{\max} = \tau_1/\tau_2$ gives a quick indication of what can be done in a particular case to improve the system, and in addition gives a good general insight into the merit of currently used corrective methods and the relationship between them.

---

[4]) See p. 352 of the book by Brown and Campbell, ref. [3]), or W. C. Elmore and M. Sands, Electronics, McGraw-Hill, New York 1949, p. 139.

Summary. For a control system to be stable the open loop gain $|KG(\mathrm{j}\omega)|$ must be less than unity at the frequency where the phase shift φ is $-180°$. In practice the control engineer adopts a safety margin of 45° and attempts to make $|KG| < 1$ when $\varphi = -135°$. In control loops where all the elements, apart from a linear amplifier, have a transfer function $(1 + \mathrm{j}\omega\tau_n)^{-1}$, and the time constants $\tau_1 \gg \tau_2 \gg \tau_3$ etc., this requirement is found to be fulfilled if $|KG(0)| \leqq \tau_1/\tau_2$. The formula is also applicable in cases where the above conditions are not entirely met. All known methods of improving a control system, viz: changing one of the time constants, using a phase-lag network or a phase-lead network, integrating element (with $\tau_{\mathrm{i}} = \tau_1$) or differentiating element (with $\tau_{\mathrm{d}} = \tau_2$), are shown to boil down to increasing τ_1/τ_2. If $|KG(0)|$ is made equal to τ_1/τ_2, the response time of the controller is roughly equal to τ_2 and independent of τ_1.

AN ELECTRICALLY SCREENED ROOM FOR MICROWAVE EXPERIMENTS

621.317.2:621.319.74

Experimental work in radio engineering is often hampered by interference caused by radiation from installations in other parts of the laboratory. This can be avoided by doing the experiments in a space completely enclosed by a conducting screen (metal sheet or gauze), which strongly attenuates extraneous electromagnetic radiation, partly by reflection and partly by absorption (joule heating due to alternating currents induced in the screen).

The first constructions of this kind made in our laboratories consisted simply of double wire-netting fixed around a wooden frame, which enclosed a space of a few cubic metres [1]). For certain purposes, however, the screening provided by these "cages" proved to be inadequate, particularly in the decimetre and

---

[1]) Part of a screened room of this kind can be seen in Philips tech. Rev. **14**, 121, 1952/53 (fig. 4).

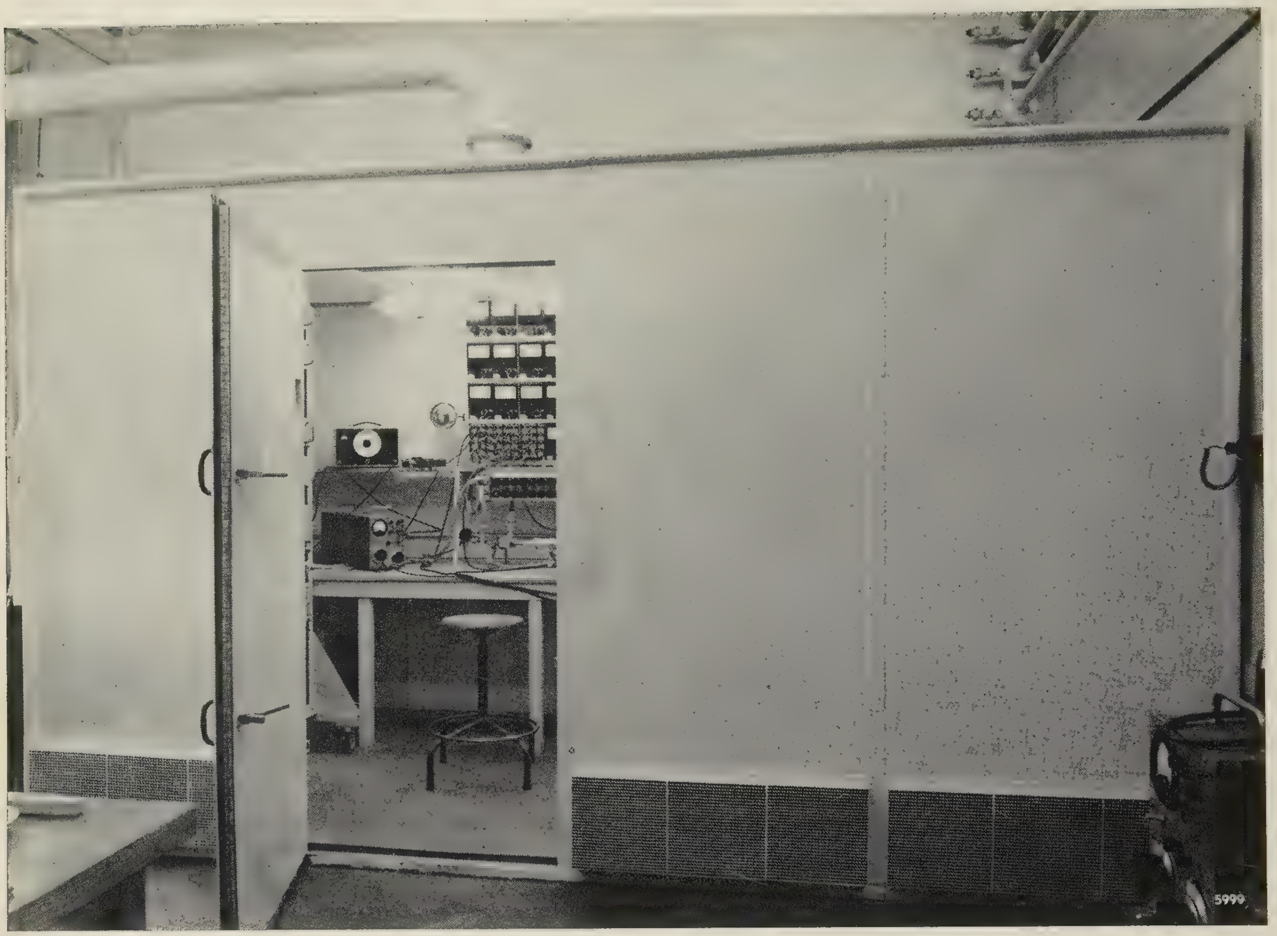


Fig. 1. General view of the electrically screened room. Floor, walls and ceiling are built up from sections each consisting of a welded angle-steel frame to which 1-mm copper sheet is screwed. The frames too are screwed together. To ensure good screening, the joints between the sections and the gaps between the door and the door-frame had to be sealed. To this end lengths of brass strip (2 mm thick, 60 mm wide) were soldered to the joints between the sections, first being screwed on in order to avoid strains during soldering. The gaps round the door were dealt with as follows. The copper sheet that forms the door was fixed to a welded angle-brass frame, around the whole periphery of which were soldered sprung strips of silver-plated phosphor-bronze (see photograph). When the door is closed, these strips press against a silver-plated copper cornice around the door frame. Good electrical connection between the cornice and the copper wall of the room is also ensured by soldering.

centimetre wavebands. A room has been in use for some time now that affords effective electrical screening up to wavelengths of about 3 cm, and is moreover designed to resemble a normal test room as much as possible. The room in question (see *fig. 1*) is 4 m long, 2 m broad and 2.30 m high; it provides ample facilities for experiments and space for three persons. The double gauze has been replaced by a single, completely closed screen of copper sheets [2]).

In such an enclosure ventilation is essential, both for the persons inside it and for cooling the equipment. The problem presented by the ventilation system, which must not impair the electrical screening, has been solved by using ventilation grids consisting basically of a number of narrow, parallel waveguides disposed at right angles to the walls. The waveguides attenuate all radiation at frequencies below a specific limit, in this case 15 000 Mc/s. The actual form of the grids is shown in *fig. 2*. They

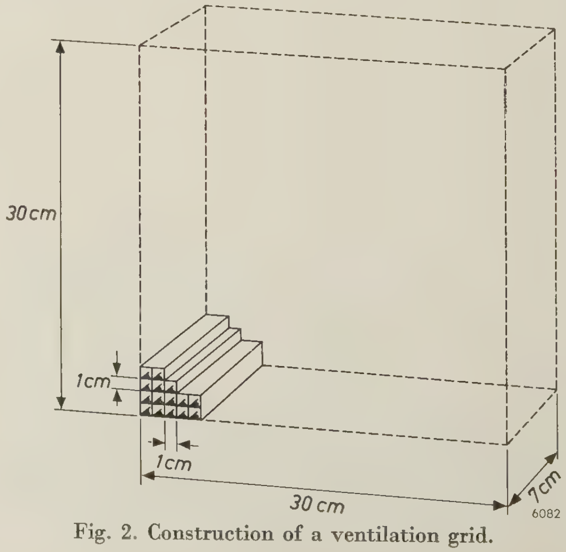


Fig. 2. Construction of a ventilation grid.

---

[2]) Double *perforated* sheets proved to be an equally effective electrical screen, but they are expensive and entail a heavy, unwieldy door of intricate structure. Moreover the walls present a disturbing background to the eye during observations (moiré patterns).

are situated near the bottom of one long and one short side wall and consist of perpendicularly intersecting copper strips 1 mm thick and 7 cm wide. The cross-section of the air channels, i.e. of the waveguides, is 1 cm². Air is extracted through a similar grid in the ceiling and an exhaust pipe. Space is left between the walls and the work benches to help distribute the upward air stream; in these gaps the supply cables are laid. As a further measure to ensure effective air distribution and to reduce air noise, a horizontal sheet of sound-absorbent material is suspended just below the ceiling.

The volume of the enclosure is 18 m³, the density of air 1.3 kg/m³. If we assume the specific heat to be 1000 J/kg °C then the mean temperature rise per watt of power dissipated in the *air* is $1/(18 \times 1.3 \times 1000)$ °C/sec $\approx 4 \times 10^{-5}$ °C/sec. If the power consumed in the room, say 2.5 kW, had to be entirely dissipated in the air — which is certainly not the case — it would be necessary, if we take a mean temperature rise of 6 °C as permissible, to refresh the air completely (i.e. all 18 m³) in a time of $6/(2500 \times 4 \times 10^{-5})$ sec $= 1$ min. The exhaust capacity of the ventilator employed is 20.5 m³/min, which is thus more than adequate.

To avoid draughts, the flow rate of the indrawn air has been put at 0.25 m/sec. The total cross-section of the inlet grids must therefore be roughly $20.5/(60 \times 0.25) = 1.37$ m². This is achieved with 15 grids measuring 30 × 30 cm², nine of which can be seen in fig. 1.

To comply with safety regulations, a person working in the screened room must be observable from outside. For this purpose the door has an observation window, of the same construction as the grids and measuring 10×10×7 cm. A convex mirror is fixed to the wall directly opposite the window, enabling anyone looking in to see practically the whole of the interior (*fig. 3*).

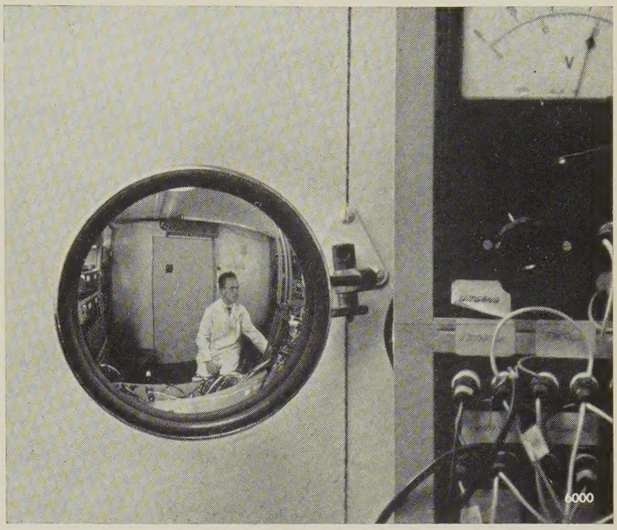


Fig. 3. Photograph taken through one of the openings in the observation window in the door. A convex mirror (see fig. 1) gives an overall picture of the interior. The window is seen in the mirror as a small dark block in the door.

In order to prevent interfering energy entering the room through the supply cables, the latter are provided with a low-pass filter consisting of a network filter combined with two coaxial filters of very simple design (see caption to *fig. 4*).

An idea of the effectiveness of the screening was obtained by setting up a transmitter outside the room, opposite those places thought likely to allow most energy to enter, i.e. the door (through gaps round the edge and through the window), the ventilation grids and the filter. For each position the

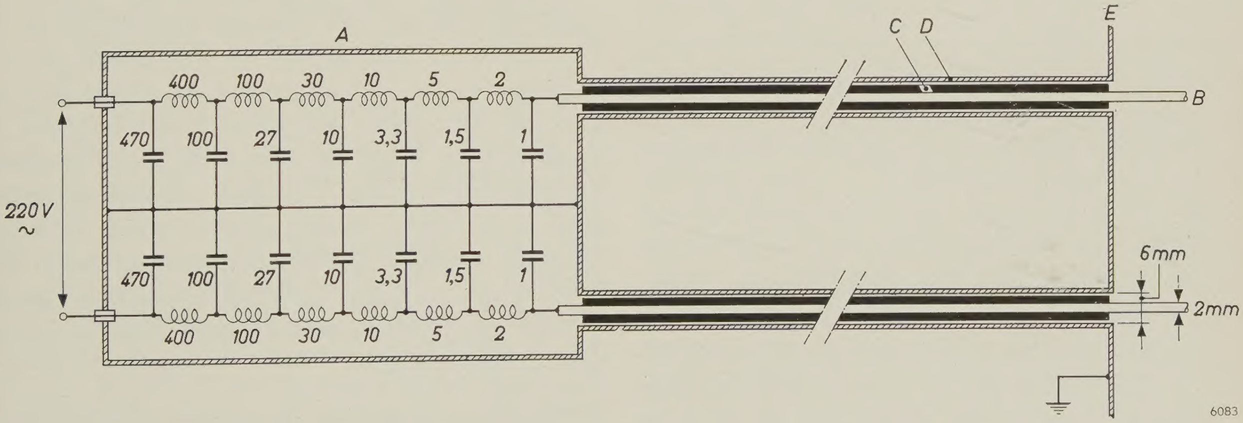


Fig. 4. Principle of the low-pass filter unit for the supply cables. At the input (left) the mains voltage is applied; *E* is the wall of the room (earthed). The brass box *A* contains a network filter for 6 A; inductances are given in μH, capacitances in nF. In the frequency range from 0.6 to 100 Mc/s the network filter gives more than 120 dB attenuation. Frequencies above 100 Mc/s, where the network filter is less effective, are attenuated by means of a coaxial filter, some decimetres long, incorporated in each of the two leads. *B* inner conductor: enamelled copper wire. *C* absorbent sheath: ferroxcube 4B. *D* outer conductor: brass pipe. A 10-cm-long test specimen of this coaxial filter gave about 20 dB attenuation at 100 Mc/s, and about 50 dB at 4000 Mc/s. The total voltage drop across the whole filter is 5 V at a load of 220 V, 6 A.

output power of a sensitive receiver inside the room measured with the door open was compared with the value found with the door shut. The results at four different wavelengths are given in the table alongside.

Table I. The degree of screening obtained.

Wavelength	Attenuation in dB		
	door	vent. grid	filter unit
300 m	> 150	> 150	> 150
3 m	> 150		
7.5 cm	45		
3 cm	25		

A. J. F. de BEER *).

---

*) Research Laboratories, Eindhoven.

Summary. Description of an electrically screened room ($4 \times 2 \times 2.30$ m) suitable for microwave experiments. Floor, walls and ceiling are constructed from 1-mm copper sheets. Ventilation is provided by clusters of air channels, 7 cm long and 1 cm^2 in cross-section, which act as waveguides with a limiting frequency of 15 000 Mc/s and thus do not impair the screening. The interior of the room can be observed from outside through a similarly designed window with the aid of a convex mirror. The penetration of interference through the supply cables is prevented by means of a network filter combined with two simple coaxial filters. Radiation from a transmitter outside the closed door is attenuated by at least 150 dB at wavelengths of 300 m and 3 m, by 45 dB at 7.5 cm, and by 25 dB at 3 cm. With the transmitter opposite a ventilation grid or the filter unit the attenuation is > 150 dB at all the wavelengths mentioned.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS BY THE STAFF OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk * can be obtained free of charge upon application to Philips Electrical Ltd., Century House, Shaftesbury Avenue, London W.C. 2, where a limited number of reprints are available for distribution.

R 404: J. B. Davies: Theoretical study of non-reciprocal resonant isolators (Philips Res. Repts **15**, 401-432, 1960, No. 5).

Modern microwave techniques make frequent use of the directional isolator, an element which attenuates the microwaves very little in one direction, but very strongly in the opposite direction. The resonance isolator consists in principle of a thin strip of ferrite placed in the longitudinal direction in the waveguide with a magnetic field across it adjusted to give resonance. Although various usable models of this isolator have already been developed, the mathematical treatment of its behaviour is still incomplete. In this article, the attenuation of the isolator is calculated for various shapes of the ferrite, using a perturbation theory which is only valid for thin strips of ferrite. It is found that the optimum shape of the ferrite is not, as was previously thought, a vanishingly thin strip, but a rod of elliptical cross-section. The results of these calculations emphasize the fact that a knowledge of the dielectric losses in the ferrite is of great importance for the determination of its optimum shape. Good results have also been obtained in practice with a thin strip of ferrite placed next to a strip of a material with a high dielectric constant. The behaviour of this combination is also calculated to a first approximation.

R 405: S. van Houten: Thermal conductivity in $\mathrm{Li}_x\mathrm{Ni}_{(1-x)}\mathrm{O}$ (Philips Res. Repts **15**, 433-436, 1960, No. 5).

Measurements of the thermal conductivity of semiconductors of the formula $\mathrm{Li}_x\mathrm{Ni}_{(1-x)}\mathrm{O}$ as a function of the lithium content. The results can be explained on the basis of the impurity scattering of phonons. The mean free path of the phonons is found to be a linear function of the mean distance between neighbouring Li atoms. This is in agreement with Kittel's theory of glasses, but not with Klemens' equation for the scattering at point defects. It is also found that this material is not so suitable for thermo-electric generators as other authors had thought.

R 406: M. L. Verheijke: The carbon content of low carbon martensite (Philips Res. Repts **15**, 437-444, 1960, No. 5).

The carbon content of low-carbon martensite can be determined dilatometrically or by means of resistivity measurements. Both methods are based on the assumption that the decreases in length and resistivity due to the first stage of tempering are linear functions of the carbon content. Experiments show that this is an acceptable assumption.

R 407: H. J. Heijn: Representations of switching functions and their application to computers (Philips Res. Repts **15**, 448-491, 1960, No. 5).

Continued from **R 399.**

R 408: D. J. Kroon: Line shape of proton magnetic resonance in paramagnetic solids (Philips Res. Repts **15**, 501-583, 1960, No. 6).

The method of nuclear magnetic resonance is being increasingly used in solid-state research of recent years. This technique has proved particularly useful in the determination of the positions of light nuclei in a crystal lattice. The information about the structure of a polycrystalline substance is mainly derived from the second moment of the resonance line.

In this thesis (Amsterdam, June 1960) an investigation of the effect of paramagnetic ions in a crystal on the second moment of the resonance line is described. The shape of the resonance line is also studied.

On the assumption that the paramagnetic ions may be regarded as point dipoles, broadening of the resonance lines may be analysed into three components:

a) composition broadening, due to the non-identical surroundings of nuclei in different unit cells of the specimen,

b) broadening due to the non-cubic surroundings of the nuclei,

c) broadening due to the fact that the crystallites constituting the porous specimen are not spherical.

The contributions to the second moment from these three sources are calculated. In order to test the theory, experiments were carried out on samples of the composition $NH_4Fe_xAl_{1-x}(SO_4)_2$. The proton magnetic resonance lines of these samples were determined at 77 °K. The variation of the second moment as a function of the iron content is in good agreement with the theory developed in this thesis. An extra broadening found especially at low iron concentrations is ascribed to a short spin-lattice relaxation time. It is found possible to indicate the positions of the protons in the crystal. It is found that even at low temperatures (20 °K) the protons are in rapid motion. The contribution of the anisotropy of the crystal to the second moment is calculated on the basis of this fact. The calculated value is of the same order of magnitude as the measured value, and the difference is probably due to the finite extent of the magnetic moment of the paramagnetic ions. The broadening of the resonance line due to the form of the crystallites was experimentally determined by some measurements on diluted samples. Finally, the shape of the resonance line was calculated and compared with the measured shape. The good agreement between the observed and calculated shaped supports the assumption made concerning the positions of the dipoles.

In the experimental part of this thesis, a simple spectrometer for nuclear magnetic resonance measurements on solids is described. In this connection, a rapid and accurate method for the design of permanent magnets is mentioned in some detail. The distortion of the resonance lines due to the method of measurement is also calculated, and means of correcting the measurements for this distortion are discussed.

R 409: J. J. Scheer: Some preliminary experiments concerning the influence of band bending on photo-electric emission (Philips Res. Repts **15**, 584-586, 1960, No. 6).

The theory that *p*-type semiconductors should exhibit a higher photo-electric quantum efficiency than *n*-type semiconductors is verified by measurements of the photo-emission from cesium-covered silicon surfaces.

R 410: G. E. G. Hardeman: Nuclear dynamic polarization in irradiated polytetrafluorethylene (Philips Res. Repts **15**, 587-597, 1960, No. 6).

When a magnetic interaction exists between the electrons and the nuclei in a substance placed in a constant magnetic field, the magnetic polarization of the nuclei can be increased by means of radiation of a suitable frequency. The extent of this "dynamic" polarization of the nuclear moments can be determined from the intensity of the nuclear magnetic resonance produced by other radiation, whose frequency is equal to the resonance frequency of the nuclei. This paper describes an investigation of the dynamic polarization of ^{19}F nuclei in samples of polytetrafluoroethylene which has been previously irradiated with fast electrons in order to form paramagnetic centres. The dynamic polarization is found to be a function of the concentration of paramagnetic centres, the intensity of the radiation, and the line shape of the paramagnetic resonance. It spreads outwards from the paramagnetic centres, which are present in much smaller numbers than the nuclei, throughout the whole spin system.

A 26: A. Klopfer, S. Garbe and W. Schmidt: Residual gases in vacuum systems (6th nat. Symp. Vacuum Technol. Trans., Philadelphia, October 7-9, 1959, Ed. C. R. Meissner, pp. 27-33, Pergamon, Oxford 1960).

Investigations of the gas composition in sealed-off high and ultra-high vacuum systems with the aid of the omegatron have shown that the gas evolution of the materials and interactions between the gases and the surfaces determine the kind of gases which are present. Such interactions are: absorption, chemical and exchange reactions. The results obtained on radio tubes, television picture tubes and systems consisting of glass and metal only are given in this paper. It appears possible to classify the composition of residual gases on the basis of the conditions prevailing in the systems in question.

A 27: K. J. Planker and E. Kauer: Bestimmung der effektiven Masse freier Ladungsträger in Halbleitern aus der Ultrarotabsorption (Z. angew. Phys. **12**, 425-432, 1960, No. 9). (Determination of the effective mass of free charge carriers in semiconductors from the infrared absorption; in German.)

The possibility of determining the effective mass of free charge carriers in semiconductors by means of infrared measurements is discussed. The main points of the existing theories of absorption are discussed, and their results are expressed in a common mathematical form. Comparison with experiment shows that the slope of all experimentally determined absorption curves is correctly predicted by Schmidt's theory of thermal scattering and Meyer's theory of scattering by defects; the absolute value of the absorption coefficients cannot however be calculated very exactly.

Absorption measurements on n-CdTe give an effective mass of about 0.24 of the free-electron mass.

A 28: H. G. Reik, H. Risken and G. Finger: Theory of hot-electron effects in many-valley semiconductors in the region of high electric field (Phys. Rev. Letters **5**, 423-425, 1960, No. 9).

The quantitative description of hot-electron effects in many-valley semiconductors is obtained by means of the solution of a properly formulated Boltzmann equation for high fields. In this formulation the actual band structure and the various scattering mechanisms are taken into account. The theory can qualitatively account for the experimentally observed anisotropy of hot electrons in a Sasaki-type experiment. For quantitative agreement intervalley scattering has to be taken into account.

A 29: S. Garbe, A. Klopfer and W. Schmidt: Some reactions of water in electron tubes (Vacuum **10**, 81-85, 1960, No. 1/2).

Traces of water which are not removed during the degassing of vacuum tubes will later be evolved, especially from the glass walls. It has been shown that this does not in fact lead to a great increase in the partial pressure of water vapour during the operation of the tube: most of the water reacts with the hot cathode or the getter mirror, forming principally hydrogen, methane and carbon dioxide.

A 30: A. Klopfer and W. Ermrich: Properties of a small titanium-ion pump (Vacuum **10**, 128-132, 1960, No. 1/2).

A small getter-ion pump is described which has a capacity of several torr-litres for chemically active gases. The maximum pumping speed for CO is about 50 l/sec. The lowest pressure obtained with this pump is less than 10^{-10} torr. The composition of the residual gases during the process of evacuation has been measured. The factors determining the lowest pressure that can be obtained are discussed. (See also Philips tech. Rev. **22**, 260, 1960/61, No. 8.)

A 31: R. Groth: Über die Temperaturabhängigkeit der kurzwelligen Ausläuferabsorption von MgO im Ultraroten (Ann. Physik **6**, 328-344, 1960, No. 5/6). (The infra-red absorption of MgO on the short-wave tail of the infra-red band as a function of temperature; in German.)

The absorption of MgO single crystals was measured for wavelengths between 5 and 10 μ (i.e. on the short-wave tail of the infra-red absorption band) at temperatures between 20 and 1500 °C. Two heaters were designed to cover the entire temperature range. The absorption coefficient at a given wavelength increases linearly with the temperature at high temperatures, as predicted by Born and Huang for absorption by second-order electrical moments. The behaviour at lower temperatures suggests that both possible excitation processes contribute to the absorption.